TU Clausthal

Domain-Independent Support for Computer-Based Education of Argumentation Skills

Doctoral Thesis
(Dissertation)

to be awared the degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted by
Frank Loll
from Burgwedel

approved by the Faculty of Mathematics/Computer Science and Mechanical Engineering,
Clausthal University of Technology

Date of oral examination
February 14th, 2012

Chairperson of the Board of Examiners
Prof. Dr. Jörg P. Müller

Chief Reviewer
Prof. Dr. Niels Pinkwart

Reviewer
Prof. Dr. Andreas Harrer

Kurzfassung

Argumentation ist allgegenwärtig und die Fähigkeit zu argumentieren essentiell in nahezu allen Bereichen des Lebens — dies gilt auf privater ebenso wie auf geschäftlicher Ebene. Um so schwerwiegender ist es daher, dass viele Menschen scheitern, wenn es darum geht begründete Argumente für einen Standpunkt vorzubringen. Klassische Lehrmethoden verschärfen diese Situation weiter, da sie sich als nur bedingt effektiv herausgestellt haben. Ein Ansatz um diesem Missstand entgegenzuwirken ist die Nutzung von Computern, um den Lehr- bzw. Lernprozess zu unterstützen.

Diese Arbeit beschäftigt sich mit den zugrundeliegenden Konzepten von computergestützter Argumentation in unterschiedlichen Domänen in denen die Fähigkeit zu argumentieren von großer Bedeutung sind, darunter die Natur- und Rechtswissenschaft ebenso wie die Ethik. Aufbauend auf einem ausführlichen Literaturreview bestehender computergestützter Argumentationswerkzeuge, die zur Forschung und Lehre eingesetzt werden, werden hierbei erfolgreiche und vielversprechende Ansätze herausgestellt. Einen Schwerpunkt bildet hierbei die Untersuchung von domänen-spezifischen Unterschieden und der Umgang mit ihnen in bestehenden argumentationsunterstützenden Systemen. Zusätzlich dazu werden die Ergebnisse einer Umfrage unter Argumentationsexperten genutzt, um die Beweggründe hinter Designentscheidungen ebenso aufzudecken wie ungelöste Probleme in diesem Forschungsbereich.

Aufbauend auf den Ergebnissen von Literaturreview und Umfrage, wird ein Architekturvorschlag für ein generalisiertes Argumentationsframework namens LASAD entwickelt. Dieses Argumentationsframework ist in der Lage identifizierte Probleme in der Lehre und Forschung auf dem Gebiet der computergestützten Argumentationswerkzeuge zu lösen. Parallel zur Referenzimplementierung wird außerdem ein Autorenwerkzeug vorgestellt und evaluiert, das es auch unerfahrenen Nutzern ermöglicht von den Flexibilisierungsmechanismen des Frameworks zu profitieren.

Abschließend wird der domänen-unabhängige Ansatz des Argumentationsframeworks auf zwei Ebenen gezeigt. Einerseits wird der bisher unklare Einfluss von Ontologie und kooperativer Arbeit auf das Ergebnis von computergestützter Argumentation im Rahmen einer kontrollierten Laborstudie untersucht. Andererseits werden externe Anwendungen von LASAD in der Lehre und Forschung zusammengefasst.

Abstract

Argumentation is an essential skill in many aspects of life - in private as well as in business. Nevertheless, many people struggle to engage in reasonable argumentation. Classic teaching methods fail to promote argumentation skills and, hence, independent knowledge acquisition by means of argumentation. An approach to deal with this issue is the use of computer tools to scaffold the education of argumentation abilities.

This thesis evaluates the underlying concepts of computer-supported argumentation in various domains that are highly dependent on argumentation such as science, the law and ethics. Based on an extensive review of existing computer-based approaches to teach argumentation skills, successful and promising concepts are identified. In addition, potential domain-specific differences in argumentation and how existing tools deal with them are highlighted. The review is extended by the results of a survey among argumentation experts in order to get further insights into the motivation and potential problems of existing approaches.

Together, this motivates the design of a generic argumentation framework called LASAD which is capable of dealing with existing problems in computer-based argumentation teaching and research. In order to enable even inexperienced users to benefit from the framework's flexibility an authoring tool is introduced and evaluated.

Finally, this thesis provides evidence for the suitability of the general concept following a two-level approach. On the one hand, this thesis evaluates the yet unclear role of ontology and collaboration on the outcomes of computer-support argumentation by means of the framework. On the other hand, the thesis closes with a summary of external applications of the framework in research and teaching.

Contents

List of Figures

List of Tables

List of Abbreviations

Notation	Description
AI	Artificial Intelligence
AIED	Artificial Intelligence in Education
ANOVA	Analysis of Variance
CMC	Computer-Mediated Communication
COMMA	Conference on Computational Models of Argument
CSCL	Computer-Supported Collaborative Learning
FR	Functional Requirement
GWT	Google Web Toolkit
IJAIED	International Journal of Artificial Intelligence in Education
ijCSCL	International Journal of Computer-Supported Collaborative Learning
IRC	Internet Relay Chat
ISSA	International Society for the Study of Argumentation
ITS	Intelligent Tutoring System
LASAD	Learning to Argue: Generalized Support Across Domains
LSAT	Law School Admission Test
NFR	Non-Functional Requirement
NoSQL	Not only SQL
RIA	Rich Internet Application
RMI	Remote Method Invocation
SUS	System Usability Scale

1 Introduction

Argumentation is an essential skill in many aspects of life - in private as well as in business. Beginning in early childhood, children learn that the successful application of argumentation skills is beneficial. Whereas babies cry to get what they want, the method of persuading is required to get more sophisticated as they grow older. Crying, which works for babies, is no longer appropriate for adolescents and adults anymore as they are expected to provide reasons for their wishes. These demands on arguers as well as the importance to argue adequately keep growing with age. In business, for instance if one tries to sell a product, the ability to argue may decide between success and failure — this is transferable to various areas and situations.

However, people often struggle to engage in reasonable argumentation (Kuhn, 1991). Thus, *"argumentation as a research topic is receiving increased attention"* (Jonassen & Kim, 2010, p. 454). The reasons for the lack of argumentation skills can be found in their education. In schools, there is, in most cases, no explicit teaching of argumentation abilities present. Subjects and domains that are predestinated to teach argumentation (e.g., science) do not promote the education of argumentation skills (Osborne, 2010). Even though there are exceptions, for instance the education of law students, these attempts are not fruitful per se. Instead, a large scale evaluation with 6542 students of classic theoretic courses to teach fundamentals in physics revealed only an average gain of 23% between pre- and post-test scores, compared to modern teaching approaches involving student group discussions, that is interactive-engagement, which seem to be more promising showing a gain of 48% (Hake, 1998). These results are transferable to argumentation. However, the involvement of discussions and reflection among students is limited by teacher's time and availability and, hence, not feasible in typical learning environments such as schools and universities.

To solve this issue, there are attempts to provide computer-based support for teaching domain-specific skills as well as argumentation abilities. The multidisciplinary field of Computer-Supported Collaborative Learning (CSCL) emerged. It is understood as *"a method of supporting people in learning together effectively via networked computers"* (Lonchamp, 2010, p. 1). One *"flash theme"* (Stahl et al., 2007, p. 127) of CSCL is computer-based (collaborative) argumentation. Here, the computer takes the role of a mediator between arguers. By means of a shared visualization (e.g., via graphs, tables or threads) of usually abstract argument structures, the computer enables users to discuss topics based on a common ground. This discussion is meant to scaffold learning.

1.1 Motivation

However, the possible benefits of these systems come in hand with a couple of problems depending on the user group and the domain in which argumentation takes place. The first group, researchers, are interested in evaluating the effect of different system factors such as the visualization, the set of elements available to model the argument, the size of the learning groups, etc. This is problematic as there is a huge amount of different systems available, which have been developed for various, but limited and oftentimes (domain-)specific purposes. Thus, it is not surprising that they differ usually in multiple factors, which makes research in this area hardly comparable.

Directly dependent on the researchers, are the other groups — teachers and students. They must be convinced to break with traditional teaching and learning methods as this is going to result in additional efforts on their sides. Once this hurdle is cleared, the teachers are faced with problems that are typical for classroom settings such as restricted access to computers, network filters, etc. In addition, they usually have a different view on the role of the system than the developer. Thus, both sides may get disappointed by software limitations that do require compromises. Students, on the other hand, do not want to deal with issues that are common to software prototypes. This is problematic, because argumentation systems are still mainly designed for research purposes which means that the installation of these tools usually involves a rather complex process and bugs may occur. In addition, typical research prototypes are not designed to deal with alternating requirements which are likely to occur in practice.

Although this description can only give a short impression of the current situation, it is obvious that these problems will hinder the successful use of software designed to support argumentation — in school as well as in science. Whereas the involvement of multiple groups raises a lot of problems already, it is getting worse when considering the variety of domain-specific differences.

1.2 Research Questions

In this thesis, I will investigate these and other problems that frequently occur in computer-supported argumentation as well as the contexts in which they occur. This way, I will identify potential points for improvements. However, CSCL is a research field involving researchers from multiple disciplines including pedagogy, sociology, psychology, and computer science. Unsurprising, each field has different experiences and views on the field which directly influences the expectations and foci of interest (cf. Lonchamp (2010)). In this thesis, I will focus mainly on the computer scientist's view on the education of argumentation abilities and how they can be applied to learn domain knowledge. Therefore, The central question that will be evaluated in this thesis is:

How can computers provide domain-independent support for the education of argumentation skills?

This central question, however, will split up into the following parts:

1. *What are the differences and similarities of argumentation in different domains and how may these influence the design of tools that support argumentation?*
2. *Which practices have shown to be effective in existing approaches to scaffold argumentation, which failed and what are the reasons for it?*
3. *How can the divergent requirements and expectations of various domains and multiple groups be met by a single framework?*
4. *Is the proposed framework an appropriate means to conduct more comparable research of open issues in computer-supported argumentation?*
5. *How can the use of computer-based argumentation tools be facilitated to promote their actual use in education?*

1.3 Method

To answer these questions, I will, as a first step, identify differences and similarities of various argumentation domains exemplified by argumentation in the domains of law, ethics, and science. By means of examples of existing tools, I will show how the domain influenced the design of these systems. In a detailed review of existing argumentation systems, I will extract similarities that are common to most domains and, in particular, I will identify practices that have shown to be effective in the education of argumentation skills. This will provide answers to research question 1 (*"What are the similarities and differences of argumentation in different domains and how may these influence the design of tools that support argumentation?"*).

A literature and system review covers past developments. Nevertheless, this is only one side of the coin as it does neither cover current developments nor the motivation behind design decisions. Thus, I will conduct a survey among experts in the area of computer-based argumentation including researchers, teachers and developers of argumentation systems to fill the gap between existing approaches and current developments. Together with the review of existing approaches to support argumentation, this will answer research question 2 (*"Which practices have shown to be effective in existing approaches to scaffold argumentation, which failed and what are the reasons for it?"*).

Based on the results of the review and the survey, I will compile a list of requirements of a generic argumentation system that is able to deal with domain-specific differences in argumentation. The elaborated requirements will motivate the architectural design of a generic framework that can be applied to various domains. As proof of concept, I will rebuild existing argumentation systems from multiple domains by means of the proposed framework. This way, I will provide an answer to research question 3 (*"How can the divergent requirements and expectations of various domains and multiple groups be met by a single framework?"*).

As a next step in order to answer research question 4 (*"Is the proposed framework an appropriate means to conduct more comparable research of open issues in computer-supported argumentation?"*) I will show that the framework is not only able to emulate existing approaches, but it is also able to be used for further research in the area of argumentation. Therefore, I will present a study that serves multiple purposes: (1) It will show that the framework is able to address open issues in the area of computer-based argumentation. In particular, I plan to evaluate the impact of ontology and collaboration on the outcomes of computer-based argumentation. These factors have not been fully explored, yet. (2) It will prove that the framework's underlying flexible approach is adequate to reduce development efforts to a minimum and that it is able to deal with alternating requirements which are common to everyday use.

Finally, I will propose an authoring tool that is able to completely eliminate the need for code-level developments to make use of the framework's flexibility. This way, I hope to encourage teachers to use such systems in the classroom. To highlight the simplicity of the system's use, I will present a second study which tests the suitability of the developed authoring tool to facilitate the system's configuration to concrete needs in practice among teachers and researchers.

In addition, I will point to external studies and application scenarios in which the framework has been used successfully. This way, I will close research questions 5 (*"How can the use of computer-based argumentation tools be facilitated to promote their actual use in education?"*). By means of short descriptions of the study conditions, methods and results, I will highlight the framework's capability for practical use.

2 Argumentation in Theory and Practice

2.1 Definition: What Is Argumentation?

As a first step, it is important to clarify what is meant by *argumentation*. Therefore, I want to start with the following definitions that can be found in the literature:

> "*[...] an argument is regarded [...] as a dialogue between two (or more) people who hold opposing views. Each offers justification for his or her own view, and [...] each attempts to rebut the other's view by means of counterargument.*" (Kuhn, 1993, p. 322)

> "*[...] the goal of argumentation is to persuade or convince others that one's reasoning is more valid or appropriate*" (Carr, 2003, p. 76)

> "*Argumentation [...] is the production of opinions accompanied by reasons in favor or against, in combination with questioning, clarification, explanation and acknowledgment*" (Munneke et al., 2003, p. 115)

In these definitions, it becomes clear that argumentation involves a complex process which can be summarized as *"reasoning"* between multiple persons or parties. In this thesis, these parties are expected to be humans. Other perceptions of argumentation, for instance in the field of Artificial Intelligence (AI) (Rahwan & Simari, 2007), will not be covered. The reasoning consists of different activities including *"questioning, clarification, explanation and acknowledgment"* (Munneke et al., 2003) as well as *"offer justifications"* (Kuhn, 1991). However, the goal of the argumentation process is not clearly defined. Here the definitions of Kuhn (1991) and Carr (2003) differ in an important aspect from the one of Munneke et al. (2003). The former assume the existence of (opposing) opinions. Therefore, the goal is to advance or defend these opinions against others. The latter does not make this assumption. Instead, Munneke et al. (2003) describe the *"production of opinions"* by means of the argumentation process in contrast to advancing or defending an existing opinion as goal of argumentation.

This differentiation is further elaborated by Andriessen (2005), who distinguished two types of argumentation: (1) aggressive argumentation, and (2) collaborative argumentation. The first type of argumentation (aggressive argumentation) can be found, according to Andriessen (2005), "*on talk shows and in the political sphere, where representatives of two opposed viewpoints spout talking points at each other*" (p. 443). The goal of this type of argumentation is to win an argumentative fight without consideration of other points-of-view. Thus, the educational perspective which is inherent in collaborative or *productive*[1] argumentation is, at least in most parts, absent. Whereas children typically engage in aggressive argumentation (cf. Andriessen (2005)), adolescents are expected to engage in productive argumentation.

However, the definitions of Kuhn (1991) and Carr (2003) are not per se aggressive. As a matter of fact, they do not mention an important aspect in their definition: Are other opinions considered? This consideration of other perspectives can occur in different ways. On the one hand, other opinions may be considered including their pros and cons. Nevertheless, the own opinion may still be favored afterwards, because the reasons are more valid. Carr (2003) refers to this point as *"convinc[ing] others"*. Combining this view with the goal described by Munneke et al. (2003), the concrete result of argumentation is accepted knowledge. Here, a typical example is science. By means of observations, usually multiple explanations for certain phenomena are developed. By argumentation, the most convincing one is chosen while the others will be dropped. The accepted theory will then hold until new observations or data lead to a revision. On the other hand, Carr (2003) mentions persuasion. In this case the other opinion may be considered, but only one-sided, that is just the argumentative lacks are used to strengthen one's own position. In contrast to the first approach, the alternative opinion is never considered to be acknowledged. A typical example can be found in legal argumentation where both parties, that is the lawyer of the defendant as well as the lawyer of the plaintiff, intend to persuade the judge or jury of their perspective and, thus, win the case for their party.

In this thesis, the educational perspective is central. Strict aggressive argumentation is usually not educational and, hence, not considered here. The focus of the remainder of this work will be on productive argumentation. Nevertheless, semi-aggressive argumentation as, for instance, present in legal argumentation (cf. previous example) will also be evaluated because it still has an educational value which is to evaluate others' arguments in order to find potential weaknesses. As mentioned before, argumentation typically involves multiple parties.

2.2 Educational Perspectives of Argumentation

The educational perspective of argumentation can serve two purposes and these perspectives differ in their goals. The first one is the *learning to argue* perspective. Here, the education of skills which are involved in argumentation is central. The second one is the *arguing to learn* perspective. In this case, the application of argumentation abilities is used to learn domain knowledge. Both perspectives are present in the literature (e.g., Andriessen et al. (2003); Andriessen (2005); von Aufschnaiter et al. (2008); Jonassen & Kim (2010); Osborne (2010)). In the following, these two perspectives and the resulting implications for teaching will be elaborated further.

[1] Instead of using the term *collaborative* argumentation as Andriessen (2005) does, I prefer the term *productive* argumentation. The reason for that is that even aggressive argumentation can take place collaboratively when parties arguing against each other instead of individuals

2.2.1 Learning to Argue

In the previous section it became clear that argumentation is a complex process involving multiple activities. From the learning to argue perspective (von Aufschnaiter et al., 2008; Jonassen & Kim, 2010), the goal is to promote skills which are essential for the activities involved in the argumentation process. According to Kuhn (1991), these skills include the following:

1. The skill to generate causal theories in order to support claims (Kuhn, 1991, Chapter 2, Causal theories)
2. The skill to provide evidence to support the generated theories (Kuhn, 1991, Chapter 3, Evidence to support theories)
3. The skill to generate alternative theories (Kuhn, 1991, Chapter 4, Alternative theories)
4. The skill to imagine and discuss counterarguments to the existing theories (Kuhn, 1991, Chapter 5, Counterarguments)
5. The skill to rebut alternative theories (Kuhn, 1991, Chapter 6, Rebuttals)

Thus, an argument is considered good once these parts are incorporated. When considering this enumeration as basic argumentation process, most people fail already in the second step, that is the provision of well-founded reasons for their claims (Jonassen & Kim, 2010). Further, the *"most common weaknesses in argumentation is the lack of counterargumentation"* (Jonassen & Kim, 2010, p. 442). People are usually biased towards their own position, that is they avoid or even refute to provide reasons to support alternative theories as shown for instance by Zeidler (1997). In connection to the prior section, one can conclude that most people are not used to a productive argumentation style, but rather to an aggressive one.

Therefore, the educational challenge from the learning to argue perspective is to identify typical problems that occur in argumentation and finding ways to fix them in order to promote argumentation abilities.

2.2.2 Arguing to Learn

The command of argumentation skills is an essential prerequisite of the arguing to learn perspective (Andriessen et al., 2003; Andriessen, 2005; von Aufschnaiter et al., 2008; Jonassen & Kim, 2010; Osborne, 2010). In this educational perspective the role of argumentation can be understood as following:

> *"Argumentation is the means by which we rationally resolve questions, issues, and disputes and solve problems"* (Jonassen & Kim, 2010, p. 439)

A typical problem that is solved by argumentation is the active construction of knowledge (Andriessen et al., 2003), the key goal of science. In classroom, knowledge is usually taught following a passive approach and argumentation is nearly absent (Osborne, 2010). Instead, knowledge will be imparted in form of results. While these results, at least in some cases, come along with an explanation, it is usually not clear why alternative explanations and theories are considered wrong. However, *"comprehending why ideas are wrong matters as much as understanding why other ideas might be right"* (Osborne, 2010, p. 464)

This point is confirmed by a meta-analysis (Hake, 1998) of 14 (physics) classes taught either by means of classic, passive approaches or by means of (inter-)active approaches. Here, the former resulted in a significantly lower improvement (average learning gain 23%) with respect to domain knowledge than the latter (average learning gain 48%). These results can be transferred from science into other domains which apply similar learning methods such as law, ethics, philosophy, medicine, etc.

Thus, the educational challenge from the arguing to learn perspective is to provide means that support the use of argumentation for more sustainable knowledge acquisition. These means should be able to solve existing problems that are usually mentioned when asking for the reasons why interactive teaching methods are not applied, for instance the workload of teachers or the motivation of students to engage in argumentation. The benefit is obvious: Instead of just learning facts, the use of argumentation allows the learners to reconstruct knowledge at any time by interpreting background material while, at the same time, rebut potentially wrong theories.

2.3 The Ill-Defined Nature of Argumentation

While critical on the one hand, the education of argumentation skills and the use of argumentation to gain domain knowledge is a complex endeavor on the other hand. In primary school, typical problems that students are required to solve are well-defined, that is they have *"a single [...] answer and, usually, an optimum strategy for proceeding from problem presentation to solution"* (Kuhn, 1991, p. 7). An example is basic arithmetics. The calculation of $1 + 1 = ...$ has obviously a single solution and a single solution strategy. Over time even school problems get more complex. With growing complexity, the solution space, that is the number of possible solutions and solution strategies, grows as well. In the example, the solution space could be extended by using an additional numeral system such as the binary numeral system which is typically used in computer science. Thus, the solution of the task could be either 2 (in the decimal numeral system) or 10 (in the binary numeral system). However, in math most problems stay solvable even though they may be highly complex.

In contrast, there are domains which involve problems that have no definitive solution.

These domains and problems are called *ill-defined*[2]. However, there is no final definition of ill-defined. As a matter of fact, the definitions are manifold (e.g., Reitman (1964); Newell (1969); Simon (1973); Kuhn (1991); Jonassen (1997); Shin et al. (2003); Voss (2006); Lynch et al. (2006, 2009); Mitrovic & Weerasinghe (2009); Jonassen & Kim (2010)) and mostly dependent on the context. In this thesis, I will use one of the more recent definitions proposed by Lynch et al. (2009), who reviewed existing definitions and came up with the following definition of *ill-defined problems* and *ill-defined domains*:

> *"A problem is **ill-defined** when essential concepts, relations, or solution criteria are un- or under-specified, open-textured, or intractable, requiring a solver to frame or recharacterize it. This recharacterization, and the resulting solution, are subject to debate."* (Lynch et al., 2009, p. 258)

> *"**Ill-defined domains** lack a single strong domain theory uniquely specifying the essential concepts, relationships, and procedures for the domain and providing a means to validate problem solutions or cases. A solver is thus required to structure or* recharacterize *the domain when working in it. This recharacterization is subject to debate."* (Lynch et al., 2009, p. 258)

Domains which require argumentation to solve problems such as science, the law and ethics are inherently ill-defined. In addition, problems in ill-defined domains are typically ill-defined as well. In order to decide whether a domain or problem is ill-defined, Lynch et al. (2009) collected a set of criteria. These criteria are:

1. The involvement of open-textured concepts and competing domain principles that are subject to debate
2. A lack of widely accepted domain theories identifying relevant concepts and functional relations
3. The problem cannot be partitioned into independent subproblems
4. There are prior cases that are facially inconsistent
5. There is a need to reason analogically with cases and examples
6. The solution space is large and/or complex which makes it impossible to enumerate all possible characterizations or solutions
7. A lack of formal or well-accepted methods to verify solutions
8. A lack of clear criteria by which solutions are judged
9. There is no solution; instead it may be readdressed by multiple, often distinct, solutions
10. The involvement of disagreements among domain experts regarding the adequacy of the solutions

[2] In the literature the terms ill-defined and ill-structured are typically used interchangeably. However, there are exceptions which make a distinction between these terms (e.g., Jonassen (1997)). In this thesis I will use the term ill-defined.

11. Solvers are required to justify their solutions through argument

To illustrate these criteria, imagine the following problem, which has been subject to debate in various domains such as public policy, ethics and science for the last decade: *How will the global warming evolve in the next 50 years?* In order to answer the question, argumentation is used.

Therefore, the first criterion is obviously fulfilled since argumentation typically involves open-textured concepts. In addition, there are multiple domains involved, all of them applying different principles. These principles are ambivalent even within a single domain. The second criterion is directly connected to it. The theories which caused the ongoing climate differ as well. While a majority is the opinion that the CO_2 production is the main source, others doubt this. A partition of the main problem in independent subproblems is also not possible. Factors, that are clearly involved in the topic are, for instance, transportation and power supply. However, these subtopics are not independent from each other (criterion three) — just think about electric cars, which are highly dependent on power supply. The fourth criterion is also fulfilled. In fact, there are reports that tried to answer similar questions beforehand. The results are in most cases inconsistent. Since there is no safe knowledge of what the world will be in 50 years, it is obvious that multiple imaginable scenarios must be played through. Thus, the fifth criterion is fulfilled as well. With each additional scenario, the solution space grows. Since the number of potential scenario is infinite, this is true for solution space, too. Therefore, the sixth criterion is given. In addition, the problem deals with potential effects in the future, which makes it impossible to verify the solution as mentioned in criterion seven. The eighth criterion, however, is only fulfilled in parts. Whereas a potential measurement of success is the temperature in 50 years, it cannot be judged beforehand. Further, there is no solution in this case as mentioned in criterion nine. Instead, there are multiple imaginable solutions, which highly depend on the assumptions beforehand and the domain in which they are proposed. This leads directly to criterion ten, which describes a disagreement among experts regarding the adequacy of the developed solutions. Since there is no accepted method to judge the quality and there are divergent theories at the beginning, each generated solution will be subject to debate. This debate, however, will use argumentation, which is the last criterion, and at the same time the only possible method to deal with the problem.

In conclusion, this concrete problem is clearly ill-defined. Additionally, argumentation itself shares a lot of these problems and, hence, is ill-defined as well. This has major implications for the education of argumentation skills, especially when computers are involved. Since there is no obvious right or wrong way to argue and arguments are clearly open-textured, it is hard to apply automated analyses. Another direct consequence is that *"students [...] prefer their own positions rather than the perspective of others"* (Nussbaum et al., 2007, p. 485) when faced with the uncertainty caused by ill-definedness. Moreover, teachers usually assess the quality by means of structural guidelines or argumentation models such as the argumentation scheme proposed by Toulmin (2003), even though they can only be used as heuristic. Further, these argumentation models are still subject of debate, as described for instance in (Newman & Marshall, 1991). As a matter of fact, argument parts are highly dependent on the context in which they occur. Thus, even experts do not always agree on the argumentative quality of a potential solution.

However, not all argumentation domains are ill-defined. An example is mathematics in which clear rules exist and decisions between right and wrong can be made. Furthermore, the distinction between well-defined and ill-defined is not a clear cut, but rather a continuum without clear borders (Le et al., 2010) or even clear classifications. In fact, problems are more likely to appear ill-defined for novices, even though they are rather well-defined. Thus, domain-knowledge and/or experience is beneficial or even required to categorize a problem as well as to generate a potentially "good" solution for an ill-defined problem (Voss et al., 1983; Shin et al., 2003).

According to Jonassen & Kim (2010), argumentation has been shown to support learning to solve both well-defined and ill-defined problems. Nevertheless, in ill-defined domains and problems argumentation is clearly more relevant than in well-defined ones (Cho & Jonassen, 2003). Thus, the remainder of this thesis will focus on ill-defined problems.

2.4 Domain-Specific Differences in Argumentation

In addition to the complexity of making arguments caused by their ill-defined nature, it gets even more complicated when considering multiple domains. As a matter of fact, there are domain-specific differences that needs to be considered to engage in reasonable argumentation in certain areas. In this section, I want to describe characteristics typical for certain domains. In particular, I will exemplify these differences by means of scientific and ethical argumentation as well as argumentation in law.

2.4.1 Legal Argumentation

In the legal domain argumentation is a structured process involving three parties: On the one hand, the two opposing parties, the defendant and the plaintiff, and on the other hand, the neutral party covered by the judge (or jury). The lawyers of both opposing parties try to 'win' the case for their respective clients by convincing the judge or a jury with arguments. Thus, the argumentation follows an aggressive approach in which each party only argues for its position and against the opponents position. Pro arguments for the opponent's side are considered for one's own position, but not presented to the judge or jury. The ground rules for arguing in the courtroom differ between countries. In contrast to the Civil Law premise (applied in many countries in continental Europe) in which laws are encoded as statutes, in the Common Law used in England and the U.S. the law is highly reliant on "precedent cases", that is, new cases should be decided in accordance with prior similar cases. Apparently, decisions in such cases are also based on laws and statutes. The difficulty in using these for argumentation is based on their open textured nature (von der Lieth Gardner, 1987), meaning that their conditions for application are abstract, must be interpreted in the context of specific cases, and are thus prone to subjectivity. Unlike many other types of argumentation, legal argumentation features a moderator (the judge) present at all time, who has to assure that protocol and legal ground rules are correctly applied so that either the judge himself or a jury can decide the case. Once the case is decided, there may be the option to request a change of the decision in form of an appeal. If it is accepted, the decision of the court might be changed by a higher court until no appeal is granted anymore or the decision was made by the highest court. Then, the decision is final and will, in some countries, be used as precedence case for future trials.

2.4.2 Ethical Argumentation

In ethical argumentation, there is no authoritative or established and structured approach to resolve ethical problems, that is, there is no judge who decides which argument is strongest and no institutional use of stare decisis (the legal principle by which judges are obliged to obey the precedents established by prior decisions). Thus, ethical arguments are typically more free-form in style and structure. Another key distinction is that the decision-making process in ethics does not always (or even typically) involve a pre-defined number of parties: even a single ethicist may present both pro and con positions or there may be more than two parties debating. Additionally, ethics cases are not constrained to binary conclusions as compared to legal argumentation. Finally, the goal in arguing and evaluating ethical problems is (typically) not to "solve a case" but rather to learn about the ethical ramifications of various actions.

2.4.3 Scientific Argumentation

In scientific argumentation, the goal is "*to produce new knowledge of the natural world*" (Osborne, 2010, p. 463). Therefore, multiple parties propose new ideas based on interpretations of observations. A classic example here is the notion of gravity by Sir Isaac Newton (⋆4. January 1643 - †31. March 1727), who was inspired by the fall of an apple. However — an idea per se is not science. Instead science can be understood as "*refinement of everyday thinking*" (Einstein, 1954, p. 290). In modern days, scientific problems are complex and, hence, must be solved by multiple parties. Therefore, this refinement involves the discussion of possible explanatory theories for certain phenomena among multiple parties of scientist, which may propose different explanations. By means of a productive argumentation approach, these explanations will get evaluated unless the most convincing theory is accepted in the community. However, this theory may be revised in favor of a new one, once there are additional evidences or new ways to analyze existing data.

2.4.4 Discussion

As exemplified by the areas of legal, ethical and scientific argumentation, there are domain-specific differences. These differences can be distinguished into basically four categories that are connected to each other: (1) the structural degree of the argumentation, (2) the involved participants, (3) the general goal of the argumentation process as well as (4) the type of argumentation.

In first category (the structural degree), the domains differ a lot. In legal argumentation, the argumentation is highly structured through laws and rules how to apply them. In contrast, ethical argumentation follows a free-form approach without concrete rules that are required to be followed strictly. In science, there is somewhat of a semi-structured approach present. There is a set of accepted research methods (e.g., experiments or simulations), but the way how they are applied and in which context can be decided by the scientist. The basic rule here is that the results can be understood and reproduced by the community.

Concerning the second category (the involved participants), the highly structured process of legal argumentation affects directly the number of involved parties. Contrary to the other domains, there is always a fixed number of three parties with fixed roles: the plaintiff, the defendant, and the judge or jury. Whereas the plaintiff and defendant only argue for their own position and against their opponent's one, the judge or jury has to come to a decision finally. This is the opposite of argumentation in science and ethics. In the latter cases, the number and role of arguers are undefined and may even change during the argumentation process. Further, a decision is not always final.

The consequence of these differences can be seen in the third (the general goal) and fourth category (the type of argumentation). In law, each party tries to win the case. To do so, one's own position is supported with evidences, while the opponent's position is weakened. With respect to the distinction presented in chapter 2.1, argumentation in law uses a semi-aggressive argumentation style, that is arguments are only presented for one's own position even though the opponent's view is considered. At first glance, this approach is similar to science, where each party tries to collect evidences for their theory. A more in-depth view, however, reveals crucial differences: Even though, there are contrary explanation attempts, the scientists are usually no rivals such as the defendant and the plaintiff in legal argumentation. Instead, they share common values and aim for the same goal (the construction of accepted knowledge) (Andriessen, 2005). Thus, the argumentation style is a productive one. Common to ethical argumentation, not only the finally accepted theory is important, but also the hypotheses and theories that may turn out to be wrong afterwards. Contrary to a trial, the supporters of a finally rejected theory, did provide an essential part to the final result. Yet, ethical argumentation differs from science, even though it follows a productive approach as well. In the latter, a result in form of an accepted theory is still desired to build the ground for further research, which is not always the case in ethical argumentation. Finally, the judgment of which theory is most plausible is not done by a mediating role such as the judge in law, but by the scientific community as a whole.

A final overview of domain-specific differences in argumentation is presented in Table 2.1. In conclusion, the gain of knowledge is the core of argumentation in science and ethics, whereas it is not in legal argumentation.

Domain	**Structural Degree**	**Involved Parties**	**Goal**	**Type**
Law	Highly structured	Plaintiff, defendant, judge / jury	Win a case	Semi-aggressive, goal-oriented
Ethics	Unstructured	Undefined; may change during argumentation process	Learn about ethical ramifications of various actions; gain knowledge	Productive, process-oriented
Science	Semi-structured	Typically multiple parties; the number may change during argumentation process	Generate knowledge through the verification and falsification of theories and ideas	Productive, goal- and process-oriented

Table 2.1: Comparison of argumentation in different domains

3 State-of-Art: An Overview of Computer-Supported Argumentation

Now that the theoretical foundations have been summarized, the role of computer support will be discussed in this chapter. The review that will be presented in the remainder of this chapter is an extended and modified version of the one presented in (Scheuer, Loll, McLaren & Pinkwart, 2010).

The chapter starts with an overview of different approaches to support argumentation (→ chapter 3.1). The purpose of this overview is two-fold. On the one hand it will introduce the overall concept of computer-supported argumentation. Therefore examples and descriptions of well-known argumentation systems will be provided to allow even inexperienced readers to understand what it is all about. On the other hand, however, the overview is intended to give a first impression of the variety and resulting differences of tools and methods that are available to support argumentation and argumentation learning.

This variety will then, in the next parts of this chapter, motivate a more in-depth look at the differences. This in-depth review of existing approaches will cover the topics visual representation of arguments (→ chapter 3.2), interaction design (→ chapter 3.3), collaboration (→ chapter 3.4), process design (→ chapter 3.5), underlying argument model (→ chapter 3.6), analysis and feedback techniques (→ chapter 3.7), architecture and technology (→ chapter 3.8) as well as a summary of empirical studies conducted in the field of computer-supported argumentation (→ chapter 3.9).

The first part of the in-depth review, visual representation of arguments (→ chapter 3.2), deals with the question how to make abstract arguments visible. Multiple approaches to visualize argument structures will be described by means of examples and potential advantages and disadvantages will be discussed.

In the second part, interaction design (→ chapter 3.3), multiple ways to create and layout arguments within argumentation systems will be reviewed. Here, a set of argument construction methods will be presented and potential benefit and drawbacks of user-controlled and system-controlled layout mechanisms will be discussed.

The third part deals with the role of collaboration in computer-supported argumentation (→ chapter 3.4). Different collaboration settings used in existing argumentation systems will be discussed and potential issues will be highlighted.

In the fourth part, process design (→ chapter 3.5), scaffolding approaches will be reviewed. This includes the definition of combinations of different learning objectives, activities, sequences, roles and representations in order to guide the learning process, especially in collaborative settings.

The fifth part, ontology (→ chapter 3.6), will evaluate the role of certain argument models, that is, elements that are available to actually create an argument. Further, the question whether there is an agreed-upon ontology in certain domains will be investigated.

In the sixth part, (→ chapter 3.7) analysis and feedback mechanisms present in existing argumentation systems will be reviewed. This includes feedback from a teacher or a moderator as well as feedback from artificial intelligence agents. Therefore, potential problems of analysis techniques on the content level as well as on the structural level will be summarized. With respect to feedback, it will be discussed who should provide feedback to whom and when as well as which way is best to give feedback.

The seventh part deals with the technological level of computer-supported argumentation (→ chapter 3.8). Here, implementation details such as the underlying architecture and logging formats will be summarized to simplify future developments in this area.

In the final part of the in-depth review, the different approaches described before will be evaluated by means of a review of empirical studies conducted in the field of computer-supported argumentation (→ chapter 3.9). Therefore, there will be descriptions of studies that aimed to provide evidence for the suitability of different approaches. By means of these descriptions promising approaches will be highlighted.

Finally, the chapter will close with a discussion (→ chapter 3.10) of the review's results. Based on the insights that have been gained in the in-depth review of existing approaches and the results of the studies in the field, open challenges will be summarized and related to the overall goals of this thesis.

3.1 Argumentation Systems

The amount of existing computer tools available to support argumentation, so called *argumentation systems*, is impressive, indicating the recognition of the importance of argumentation and the teaching of argumentation skills. Examples include (in alphabetic order) AcademicTalk (McAlister et al., 2004), Araucaria (Reed & Rowe, 2004), Argue/ArguMed[1] (Verheij, 1998, 2003), Athena (Rolf & Magnusson, 2002), Belvedere (Suthers et al., 1995, 2001), Carneades (Gordon, 2007), Collaboratorium (Klein & Iandoli, 2008), Compendium (Buckingham Shum et al., 2006), Convince Me (Schank, 1995; Schank & Ranney, 1995), Debategraph[2], Debatepedia[3], Digalo

[1]http://www.ai.rug.nl/~verheij/aaa/argumed3.htm
[2]http://www.debategraph.org
[3]http://debatepedia.idebate.org

(Schwarz & Glassner, 2007), gIBIS (Conklin & Begeman, 1987, 1988, 1989), InterLoc (Ravenscroft et al., 2008), LARGO (Pinkwart et al., 2007), QuestMap (Carr, 2003; Buckingham Shum et al., 2006), Rationale (van Gelder, 2007), Reason!Able (van Gelder, 2002, 2003), SenseMaker (Bell, 1997; Bell & Linn, 2000) and WISE (Linn et al., 2003) to mention just a few[4]. A complete list of systems and methods that have been considered for this review is given in Appendix A. Even though all of these systems share a common purpose — the support of argumentation — they differ in multiple aspects.

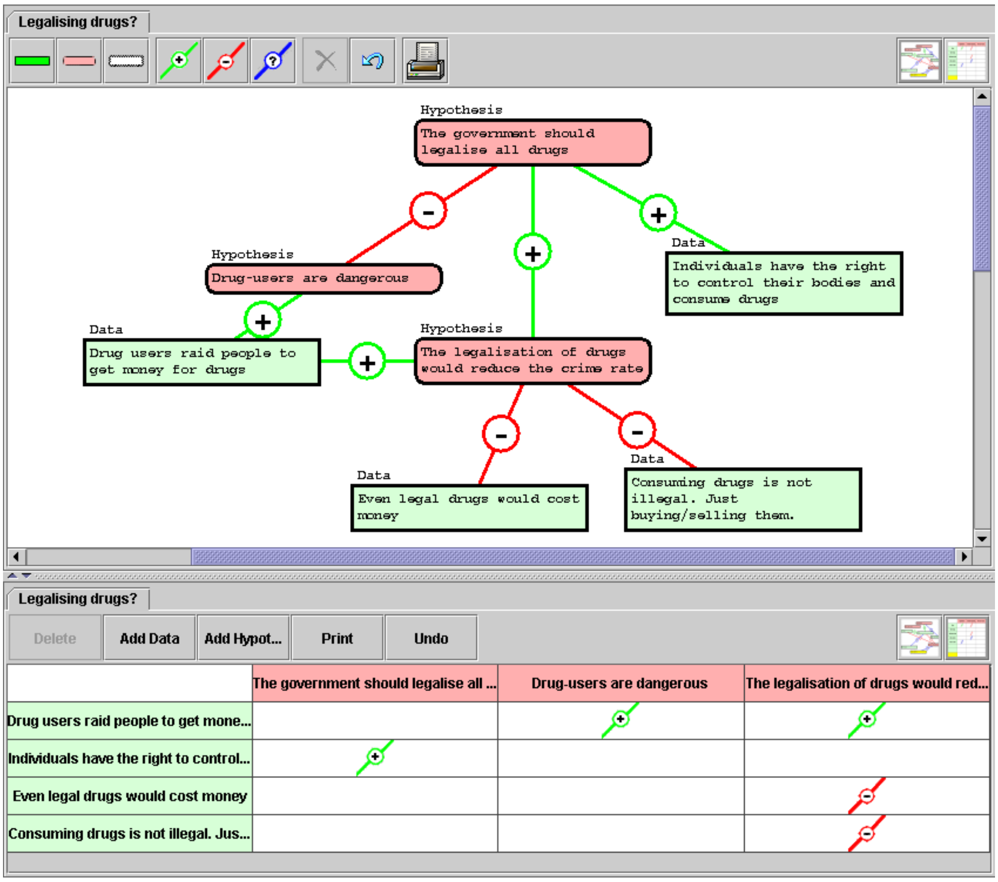


Figure 3.1: Argument representation in Belvedere 4.1

To illustrate these differences, Figure 3.1 shows Belvedere, one of the best known argumentation systems available. Belvedere was designed as graph-based diagramming

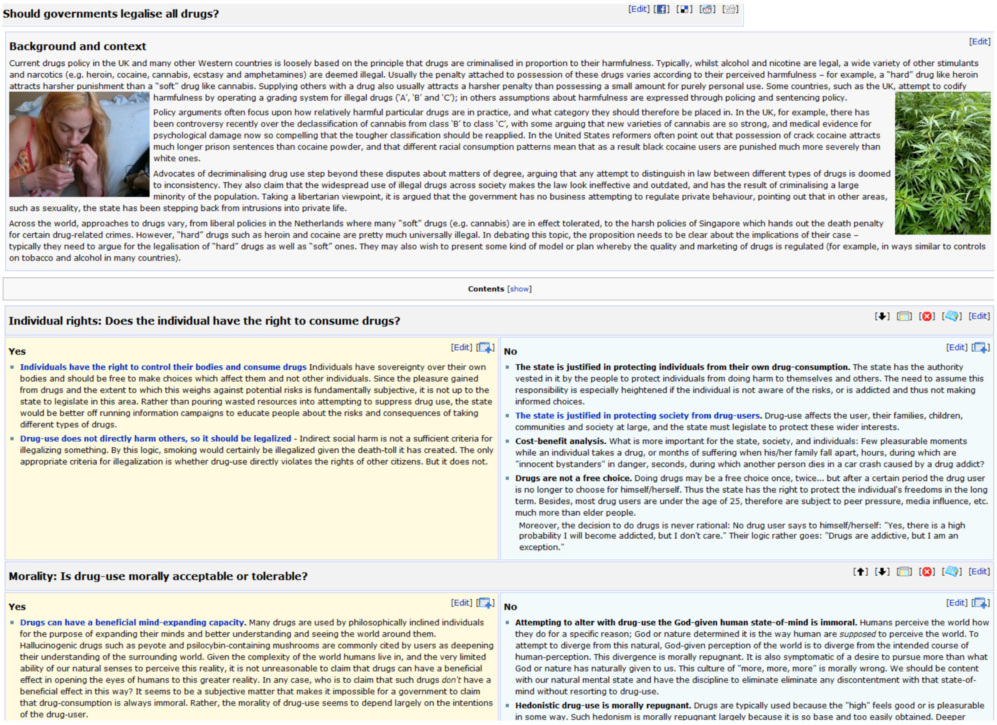
[4]This set of tools will serve as example in the remainder of this chapter. Even though these argumentation systems cover most of the features that will be discussed in this chapter, there are some exceptions. In these cases, additional tools will be mentioned. For clarity reasons, the respective literature references to the systems will only be mentioned the first time the tool is discussed.

tool for scientific argumentation. Here, the initial focus was on education of groups of secondary school children. By means of an Intelligent Tutoring System (ITS), advisory guidance was provided in order to encourage self-reflection among the students. However, in later versions the focus moved from advisory to representational guidance, that is, using specially targeted argument interfaces to bias and guide students' discourse. Alongside, complex scientific reasoning was replaced by simpler evidential argumentation involving an easier underlying argument model, which is shown in Figure 3.1. In this example, argument contributions (e.g., a data element "Drug users raid people to get money for drugs" and a hypothesis element "Drug users are dangerous") are represented as nodes, and links between the nodes represent the relationship between the contributions (e.g., "Individuals have the right to control their bodies and consume drugs" supports the hypothesis "The government should legalise all drugs").

In this description of Belvedere, there are already a lot of starting points to discuss differences among argumentation tools. The first one is the role of collaboration. As a matter of fact, quite a lot of existing systems are designed to be used individually. When considering the list above, Argue/ArguMed, Athena, Carneades, Convince Me, LARGO, Rationale, Reason!Able and SenseMaker do not provide any build-in support for collaboration. But even among the tools that support collaboration, there are differences concerning style (asynchronous versus synchronous) and dimension (small groups versus large groups) of collaboration. While Belvedere was designed to support typical class sizes, Debategraph, Debatepedia and Collaboratorium, for instance, are designed to discuss topics of global importance, such as climate change (Malone & Klein, 2007) involving large user counts.

The second difference can be found in the visualization of argument structures. In Figure 3.1, a graph-based view is presented as well as a table-based view on the same data. However, these are not the only visualizations that are available in existing tools. Even though the graph-based visualization is by far the most adopted one, there are visualizations aside of graph and tables. Here, a typical example is AcademicTalk (the predecessor of InterLoc) which makes use of a (structured) chat-style visualization of arguments. By means of sentence openers, a categorization, similar to the one applied in a graph-based visualization is achieved. Another example is the container-style / frame-style visualization of SenseMaker (as part of the KIE environment, which is the predecessor of WISE) and Debatepedia. The latter is shown in Figure 3.2. In this visualization, argument parts that belong together are in a shared frame.

Somewhat connected to the visualization is the third difference, the underlying argument model, that is the ontology. While the underlying argument model in Belvedere evolved over time from a rather complex approach to a simpler one, this is only one example. In fact, there are nearly as many different underlying argument models as there are tools available. However, this difference is, at least in parts, motivated in the different goals of the tools. Whereas an environment designed to analyze existing arguments such as Araucaria makes use of models that are especially designed for this task, e.g., the Toulmin (2003) argumentation scheme or the one proposed by Wigmore (1931), there are other environments such as Reason!Able, gIBIS, QuestMap



Should governments legalise all drugs? [Edit]

Background and context [Edit]

Current drugs policy in the UK and many other Western countries is loosely based on the principle that drugs are criminalised in proportion to their harmfulness. Typically, whilst alcohol and nicotine are legal, a wide variety of other stimulants and narcotics (e.g. heroin, cocaine, cannabis, ecstasy and amphetamines) are deemed illegal. Usually the penalty attached to possession of these drugs varies according to their perceived harmfulness – for example, a "hard" drug like heroin attracts harsher punishment than a "soft" drug like cannabis. Supplying others with a drug also usually attracts a harsher penalty than possessing a small amount for purely personal use. Some countries, such as the UK, attempt to codify harmfulness by operating a grading system for illegal drugs ('A', 'B' and 'C'); in others assumptions about harmfulness are expressed through policing and sentencing policy.

Policy arguments often focus upon how relatively harmful particular drugs are in practice, and what category they should therefore be placed in. In the UK, for example, there has been controversy recently over the declassification of cannabis from class 'B' to class 'C', with some arguing that new varieties of cannabis are so strong, and medical evidence for psychological damage now so compelling that the tougher classification should be reapplied. In the United States reformers often point out that possession of crack cocaine attracts much longer prison sentences than cocaine powder, and that different racial consumption patterns mean that as a result black cocaine users are punished much more severely than white ones.

Advocates of decriminalising drug use step beyond these disputes about matters of degree, arguing that any attempt to distinguish in law between different types of drugs is doomed to inconsistency. They also claim that the widespread use of illegal drugs across society makes the law look ineffective and outdated, and has the result of criminalising a large minority of the population. Taking a libertarian viewpoint, it is argued that the government has no business attempting to regulate private behaviour, pointing out that in other areas, such as sexuality, the state has been stepping back from intrusions into private life.

Across the world, approaches to drugs vary, from liberal policies in the Netherlands where many "soft" drugs (e.g. cannabis) are in effect tolerated, to the harsh policies of Singapore which hands out the death penalty for certain drug-related crimes. However, "hard" drugs such as heroin and cocaine are pretty much universally illegal. In debating this topic, the proposition needs to be clear about the implications of their case – typically they need to argue for the legalisation of "hard" drugs as well as "soft" ones. They may also wish to present some kind of model or plan whereby the quality and marketing of drugs is regulated (for example, in ways similar to controls on tobacco and alcohol in many countries).

Contents [show]

Individual rights: Does the individual have the right to consume drugs? [Edit]

Yes [Edit]

- Individuals have the right to control their bodies and consume drugs Individuals have sovereignty over their own bodies and should be free to make choices which affect them and not other individuals. Since the pleasure gained from drugs and the extent to which this weighs against potential risks is fundamentally subjective, it is not up to the state to legislate in this area. Rather than pouring wasted resources into attempting to suppress drug use, the state would be better off running information campaigns to educate people about the risks and consequences of taking different types of drugs.
- Drug-use does not directly harm others, so it should be legalized - Indirect social harm is not a sufficient criteria for illegalizing something. By this logic, smoking would certainly be illegalized given the death-toll it has created. The only appropriate criteria for illegalization is whether drug-use directly violates the rights of other citizens. But it does not.

No [Edit]

- **The state is justified in protecting individuals from their own drug-consumption.** The state has the authority vested in it by the people to protect individuals from doing harm to themselves and others. The need to assume this responsibility is especially heightened if the individual is not aware of the risks, or is addicted and thus not making informed choices.
- The state is justified in protecting society from drug-users. Drug-use affects the user, their families, children, communities and society at large, and the state must legislate to protect these wider interests.
- **Cost-benefit analysis.** What is more important for the state, society, and individuals: Few pleasurable moments while an individual takes a drug, or months of suffering when his/her family fall apart, hours, during which are "innocent bystanders" in danger, seconds, during which another person dies in a car crash caused by a drug addict?
- **Drugs are not a free choice.** Doing drugs may be a free choice once, twice... but after a certain period the drug user is no longer to choose for himself/herself. Thus the state has the right to protect the individual's freedoms in the long term. Besides, most drug users are under the age of 25, therefore are subject to peer pressure, media influence, etc. much more than elder people.

  Moreover, the decision to do drugs is never rational: No drug user says to himself/herself: "Yes, there is a high probability I will become addicted, but I don't care." Their logic rather goes: "Drugs are addictive, but I am an exception."

Morality: Is drug-use morally acceptable or tolerable? [Edit]

Yes [Edit]

- Drugs can have a beneficial mind-expanding capacity. Many drugs are used by philosophically inclined individuals for the purpose of expanding their minds and better understanding and seeing the world around them. Hallucinogenic drugs such as peyote and psilocybin-containing mushrooms are commonly cited by users as deepening their understanding of the surrounding world. Given the complexity of the world humans live in, and the very limited ability of our natural senses to perceive this reality, it is not unreasonable to claim that drugs can have a beneficial effect in opening the eyes of humans to this greater reality. In any case, who is to claim that such drugs *don't* have a beneficial effect in this way? It seems to be a subjective matter that makes it impossible for a government to claim that drug-consumption is always immoral. Rather, the morality of drug-use seems to depend largely on the intentions of the drug-user.

No [Edit]

- **Attempting to alter with drug-use the God-given human state-of-mind is immoral.** Humans perceive the world how they do for a specific reason; God or nature determined it is the way human are *supposed* to perceive the world. To attempt to diverge from this natural, God-given perception of the world is to diverge from the intended course of human-perception. This divergence is morally repugnant. It is also symptomatic of a desire to pursue more than what God or nature has naturally given to us. This culture of "more, more, more" is morally wrong. We should be content with our natural mental state and have the discipline to eliminate eliminate any discontentment with that state-of-mind without resorting to drug-use.
- **Hedonistic drug-use is morally repugnant.** Drugs are typically used because the "high" feels good or is pleasurable in some way. Such hedonism is morally repugnant largely because it is so base and too easily obtained. Deeper

Figure 3.2: Frame-based debate from Debatepedia: Legalization of drugs

and Compendium which aim to support a specific learning or argumentation theory (in this case the Reason! method in Reason!Able (van Gelder, 2003) and the IBIS methodology (Kunz & Rittel, 1970) in gIBIS, QuestMap and Compendium) with its ontology elements. However, even though the goal might be the same, the approaches how to achieve this goal may differ between systems. Examples are Athena, Digalo and Rationale. All of them aim to support general argumentation. Whereas Rationale uses a large set of available elements and Athena just uses a general node type for everything, Digalo makes use of a flexible argument model that can be reconfigured for each argument.

A fourth difference is the kind of support that is offered by the tool. While there are very basic tools such as Athena, which is designed to model an argument before discussing it in front of a class, there are, at the same time, tools that make use of a sophisticated analysis and feedback approach to provide live support for the argumentation process. Examples that can be mentioned here are Carneades, Convince Me, Digalo and LARGO. Carneades makes use of a formal, mathematical model to compute and assign acceptability values to propositions and supports multiple proof standards, that is procedures to derive the acceptability of a claim and associated arguments such as "preponderance of evidence". Similar decision procedures are implemented in ArguMed and Hermes (Karacapilidis & Papadias, 2001). In Convince Me, user assigned ratings to argument parts are compared to those of a computational model called ECHO. Digalo, in combination with ARGUNAUT (McLaren et al., 2010), supports human moderators in feedback provision to learners by means of observations collected and categorized via machine learning techniques. LARGO, as a last example, makes use of graph grammars in order to detect weaknesses in argument structures (e.g., hypotheses without supporting facts) created by learners. In addition, collaborative filtering is used to assess the content of these structures.

In the following sections, these and other differences will be reviewed in more detail. In addition, successful practices will be highlighted and open challenges will be summarized.

3.2 Argument Representations

A key goal of argumentation technology is to provide an external argument representation to allow users to create, manipulate, and review arguments. Argument representation formats are crucial to the interface between user and system, making different pieces of information more or less salient. System designers can exploit these properties to create user interface affordances that guide the user toward productive activity, as described in Suthers' (2003) theory of representational guidance.

A variety of different representational formats have been employed in existing systems. Some formats aim primarily at supporting communication between users while others aim to represent the conceptual structure of debate; some are used for education while others are used for collaborative decision making and deliberation; some aim at argument production, while others are used for argument analysis and evaluation. In this section, the five major representational types found in the literature will be discussed: (1) linear, (2) threaded, (3) graph-based, (4) container, and (5) matrix / table.

The simplest form of argument representation is a linear, usually textual, form. Simple computer-mediated communication (CMC) tools like chats (e.g., IRC) are used for this form of argumentation. A chat can be thought of as a written analog to (sequential) spoken communication. The main advantage of such tools is their ease of use and familiarity. A problem with chat, however, is sequential incoherence, especially in chats involving more than two or three participants (McAlister et al., 2004). This problem occurs when it is not clear which comments and responses refer to which other comments. An example illustrates this:

> *(10:01:12) - Alice: There are different kinds of drugs, for instance marihuana and hashish*
> *(10:02:22) - Bob: Marihuana is allowed in some countries*
> *(10:02:23) - John: Which is the most dangerous one?*

Here, John intends to ask Alice what is the most dangerous drug. However, because of Bob's intervening statement, this is not immediately obvious (at first glance, one could think that John wants to know which country is the most dangerous one) without a careful review of the time stamps. Threaded discussions, on the other hand, explicitly capture message-reply sequences, reducing the risk of sequential incoherence (multiple answers from different users at the same time might still be irritating), and better support users in managing large argumentation strands. While threaded discussion tools do not provide special support for argumentation, there have been attempts to support argumentation by tagging according to type, that is, annotating the role of contributions in an argument. An example for this is Hermes (Karacapilidis & Papadias, 2001), which uses a forum-like style in which contributions are marked as, for instance, issues, alternatives, and pro and con positions. The contribution categories largely differ between argumentation domains and tool purposes; this is discussed in more detail in the section 3.6.

Most argumentation tools, educational as well as general purpose ones, do not use purely textual argument representations, such as linear or threaded texts. The most frequently employed representation is a graph-style (see Figures 3.1, 3.3 & 3.4). In this approach, contributions are displayed as boxes or nodes that represent argument components, such as claims or facts. The edges (or arrows) of these graphs represent relations between the argument components (e.g., supports or refutes). There are notable differences between the kinds of argument graphs that can be created with the existing systems: Some systems use proper graphs, while others use hyper-graphs, that is, graphs that allow links between links or trees, that is graphs with a single root node. For instance, Belvedere allows a user to construct conjunctions of propositions with an "and" link, which can, yet again, be related to other nodes or links. Some systems impose no restrictions on the linking of parts, while others permit only tree structures, that is, special cycle-free graphs with a single root element (e.g., the main hypothesis or question). Graph representation systems are numerous, including Belvedere, Athena, and Digalo in the unrestricted variant and Reason!Able, Araucaria, and Carneades in the more restricted tree-based variant. The implications of these restrictions will be discussed in more depth in section 3.3.

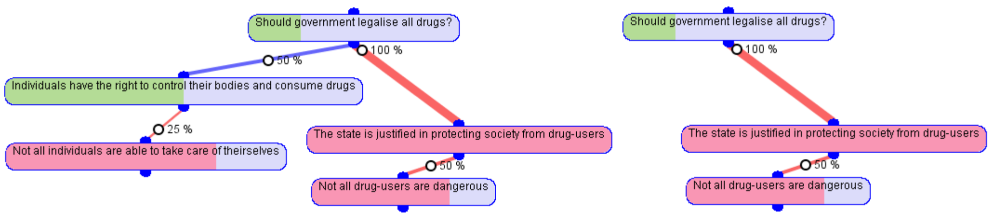


Figure 3.3: Acceptability filter in Athena (0% vs 30%)

A key asset of graphs is their explicitness and clarity, that is, each elementary knowledge unit is encapsulated as an individual node and relations between knowledge units are explicated with edges. The different types of knowledge units and relations can be easily distinguished via their visual appearance. This explicitness makes graph-style representations an excellent and intuitive knowledge modeling approach, for instance, for analyzing argument transcripts (e.g., LARGO and Araucaria) or in keeping a record of the current state of a debate to support decision making (cf. van Gelder (2003); Buckingham Shum et al. (2006)). Others use graphs as a structured medium of debate (e.g., Schwarz & Glassner (2007)) or as representational tools that provide an external and persistent stimulus to encourage and scaffold discussions, that is, students do not discuss within a graph but about a graph (Suthers, 2003). The suitability of graphs as a medium for argumentation interacts with other factors, such as the complexity of the underlying ontology and the number of discussants and synchronicity of communication. Graphs may become unwieldy when a relatively large number of participants engages in synchronous discussion using a complex ontology. Also, the user group and domain may have an impact on the suitability of graph notations: Hair (1991), for instance, reports changing the graph representation in the Legalese argumentation tool (designed to assist lawyers in constructing arguments) because lawyers strongly preferred threaded text representations.

Another approach to structuring discussions and argumentation visually is used in SenseMaker (Bell, 1997), which visualizes argumentation strands belonging together graphically via frames (windows) which serve as containers (similar to Debatepedia's representation shown in Figure 3.2). Here, each frame represents a claim that is supported by the evidence elements and other claims contained in that frame. Elements (claims and evidence) can be annotated with "usefulness scores", which are represented as colored dots (e.g., red for high). Other examples of the container visualization technique are the wiki-based Debatepedia (http://wiki.idebate.org) which provides one frame per question, containing arguments pro or con the topic (see Figure 3.2), and Room 5, a system for legal argumentation (Loui et al., 1997).

The key advantage of the containers style is the possibility of recognizing, at a glance, argument components that belong together. Yet, at the same time, this technique makes it difficult to express types of relations, because relations are expressed implicitly by one element visually residing within another element. Also, one is not able to recognize whether two components in the same frame are related to each other. Additionally, with this type of visualization, it is hard for the user to get an overview of what is happening in a large argumentation map (a similar problem to the graph style).

An attempt to visualize implicit or missing relations between argument elements is the use of a matrix argument representation in which argument components (e.g., hypotheses, facts) are the rows and columns of the matrix while the cells represent the relations between the components. This visualization method has, to my best knowledge, only been implemented in the Belvedere (v3 and v4) system (see Figure 3.1). The primary strength of a matrix in representing arguments is highlighting the missing relations between important aspects of arguments (Suthers, 2003). However, the creation of arguments via a matrix style is more abstract (and perhaps less intuitive) than, for instance, constructing a graph. Furthermore, it is not easily possible to represent links between links in the matrix style.

Some systems support multiple visualizations, that is, different views of the same argument, each of which (possibly) focuses on different aspects with different affordances. These views are sometimes provided simultaneously: In the last published version of Belvedere (v4.1), for instance, it is possible to use both a graph view and a matrix view at the same time (see Figure 3.1) and hence, to (potentially) benefit from their different strengths (see Table 3.1 for an overview). Other systems also support multiple visualizations but not simultaneously, that is, the user has to switch between the different views. Examples include Debategraph (http://www.debategraph.org), which offers a graph style and a threaded textual representation, Araucaria, which allows the user to switch between three representational notations that differ in terms of layout and ontology, and CoPe_it! (Karacapilidis et al., 2009), which offers views that differ in terms of formality and available actions to support different argumentation phases (e.g., brainstorming versus decision making).

Representation style	**Typical uses**	**Pros**	**Cons**
Linear (e.g., chat)	Discussions (especially synchronous)	Familiar and intuitive to most users, easiest to use	Risk of sequential incoherence (McAlister et al., 2004)
		Best to see temporal sequence and most recent contributions	Not suited to represent the conceptual structure of arguments
			Lack of overview
Threaded (e.g., forums, Academic Talk)	Discussions (especially synchronous); modeling	Familiar and intuitive to most users, easy to use	Moderately hard to see temporal sequence (because of multiple threads) as compared to Linear
		Easy to manage large discussions	Limited expressiveness (only tree-like structures)
		Addresses issue of sequential incoherence	
Graph (e.g., Belvedere, Digalo)	Discussions; modeling	Intuitive form of knowledge modeling (Suthers et al., 1995; van Gelder, 2003)	Hard to see temporal sequence
		Highly expressive (e.g., explicit relations)	Lack of overview in large argumentation maps (need a lot of space, can lead to "spaghetti" images (Hair, 1991; Loui et al., 1997))
		Many graph-based modeling languages exist	

Representation style	Typical uses	Pros	Cons
Container (e.g., SenseMaker, Room 5)	Modeling	Easy to see which argument components belong together and are related	Limited expressiveness (e.g., only implicit relations, only tree-like structures)
			Lack of overview in large argument maps because of missing relations
Matrix (e.g., Belvedere)	Modeling	Easy systematic investigation of relations	Limited expressiveness (e.g., supports only two element types (row, column), no relations between relations)
		Missing relations between elements are easily seen	Uncommon (non-intuitive) way of making arguments

Table 3.1: Comparison between argument visualization styles

Whereas the above approaches offer different views of the same argument, other systems provide different representational tools to support different argumentation-related activities. Some systems offer a more structured visualization style (e.g., graphs) for knowledge modeling together with more lightweight CMC facilities (e.g., chat) for communication purposes. For instance, Munneke et al. (2003) combined the use of chat and the collaborative writing of an argumentative text using the tool TC3 (Text Composer, Computer-supported, and Collaborative); Lund et al. (2007) provided a chat and graph-based tool in parallel during a discussion phase using the CSCL environment DREW (Dialogical Reasoning Educational Webtool); Suthers et al. (2008) have investigated how to integrate graphs as a conceptual knowledge representation with a chat and threaded discussion as the actual communication medium. Multiple representations can also be used in different phases of a longer argumentation task. For example, the CoChemEx system (Tsovaltzi et al., 2010) uses a more structured graphical representation for planning problem-solving activities and the interpretation of results while using a chat to coordinate and deliberate during the problem-solving activity itself. A second example is the work of Lund et al. (2007), in which students first debate in a chat (as medium) and then represent the debate as a graph. Some systems offer configuration facilities to define when tools are available during the learning process. An example is CoFFEE (Collaborative Face to Face Educational Environment Belgiorno et al. (2008)). CoFFEE provides a chat to send quick messages, a threaded forum for more structured discussions, a graphical discussion tool for collaborative brainstorming, and a collaborative writing tool.

The use of multiple representations is a challenge for data analysis in empirical studies as well as for automated, system-triggered analyses especially when students do not use the representations in expected ways. For instance, students might distribute their activities across multiple, available channels without considering the intended use of each channel. In such cases, it is necessary to integrate data from the different sources to evaluate the overall argument. In section 3.9, a number of studies dealing with the topic of using multiple representations will be described in more detail.

3.3 Interaction Design

In this section, the different interaction design techniques used in argumentation systems will be discussed. Therefore, I will describe how users can interact with argument representations and other interface widgets of argumentation systems, emphasizing the important aspects of computer-mediated human-to-human interaction and visual representations. Depending on intended purpose, different systems offer different modes of argument creation and manipulation, support individual use or collaboration, allow for synchronous or asynchronous student-to-student interaction, and provide specific features to help users.

First, I will address the creation of arguments. Arguments can be thought of as sets of interrelated components: basic, self-contained "argument moves" made by arguers (e.g., a claim, a hypothesis). There is no general consensus in demarcating elementary units, that is, where one unit ends and another begins. This decision often depends on the specific objectives of system and users (Reed & Rowe, 2004). One way to scaffold users is to provide a domain ontology, that is, categories or sentence openers that define elements of interest (see section 3.6). Here, another dimension will be reviewed, namely the amount of autonomy that the user has in deciding on the content of her argument components. This point is critical for learning applications, because different approaches may promote different learning goals such as argument production, analysis, and evaluation. The following five classes of argument construction can be identified:

1. Free-form arguments: The students (or users) are free to create elementary argument components on their own without restrictions, except for a possibly predefined topic. (Example systems: Digalo, Athena).
2. Argumentation based on background materials: Background materials are given to promote the argument. Based on the given materials, students create and interrelate elementary argument components. For instance, one version of Belvedere provides a collection of hypertext pages about scientific controversies. Similarly, students in SenseMaker use World Wide Web resources as pieces of evidence in their arguments.
3. Arguments rephrased from a transcript: Students review an already existing argument and, based on this, reconstruct the argument in a more structured form. For instance, LARGO provides a written protocol from an oral legal argument, which is analyzed by students and encoded as an "argument map". In contrast to a corpus of background materials, a transcript is a much more

focused resource that already contains the complete argument, but in a complex, unstructured form that is not easily understandable.

4. Arguments extracted from a transcript: Here, students are given a transcript to analyze, but instead of paraphrasing the existing argument, they are prompted to copy passages directly from this transcript to elementary argument elements (Example: Araucaria).

5. System-provided knowledge units: Elementary components are predefined. The student's task is to choose from the set of predefined components and connect them in an appropriate way to define how they relate to one another. (Example: Belvedere v2 (Suthers et al., 2001)).

Approach (1) clearly aims at supporting argument generation skills. Approaches (3) and (4) are typically used to train argument analysis skills. Approaches (2) and (5) support both generation and analysis skills. The different approaches to argumentation support vary in the degree of user and system control: In approach (1), users have complete control of the contribution content. In approaches (2) and (3), users are in control of the contribution content, but their work relies on given background materials or a transcript. In approach (4), users only decide on the segmentation of information; the actual textual content is taken from the transcript. Finally, in approach (5), elementary knowledge units are completely predefined for the user.

These argument approaches are not mutually exclusive: Some systems integrate several approaches. One example is Rationale, which supports freely created arguments as well as arguments extracted from a transcript (in essay planning mode). Belvedere v2 combines approaches (2) and (5), that is, students can make use of predefined knowledge units but can also write their own contributions, possibly based on given background materials. Araucaria mainly uses approach (4), but also supports the user in freely entering elements to add reconstructed information that is only implicitly contained in the transcript.

Approaches (4) and (5) have the advantage that the system "knows" the meaning of elementary knowledge units used by the student because they are predefined (approach 5) or can be easily traced back to a transcript passage that is "known" in advance (approach 4). LARGO (approach 3) achieves the same effect by letting the student explicitly define which transcript passage is encoded in a diagram element using markup techniques. Making such information accessible to the system makes it much easier to automatically analyze the produced arguments. (This will be further discussed in section 3.7.) Graphical argumentation systems sometimes restrict not only the creation but also the visual arrangement and structural relationships of contributions in the workspace. Conversely, many systems allow the user to create contributions and relate them to one another freely. Thus, the layout and structure depend on the users' preferences and choices and can carry a user-intended meaning (e.g., arguments in close proximity to one another belong together, even if they are not formally connected). This method is used, for example, in Athena, Digalo, Convince Me, LARGO, and Belvedere.

Another approach is to have the system control the layout. For instance, Reason!Able forces a Premise-Conclusion tree structure with "objecting" and "supporting" arguments as the children of the referred claim. Araucaria enforces argument layout according to Walton, Wigmore, and Toulmin diagrams. Such approaches lower the risk of cluttered arguments and ensure adherence to syntactic constraints, which is required by some automated analyses (e.g., in Carneades, see section 3.7). Of course, restrictions on layout may not be appropriate in all cases. In particular, the more restrictive, system-controlled approaches are more typically seen in formal modeling scenarios. Advantages and disadvantages of the user- versus system-controlled approaches, together with characteristics of the approaches (that can be either pro or con) are summarized in Table 3.2.

Layout control	**Characteristics**	**Pros**	**Cons**
System-controlled	System "guides" user through the construction; some diagram aspects cannot be manipulated directly by the user	Avoid unwanted syntactic constructs (e.g., cyclic arguments) Clear and consistent layout Easy to read	Too restrictive in some pedagogical scenarios (errors and weaknesses as opportunities to teach)
User-controlled	Diagram clarity and consistency depend on user's proficiency and discipline User can directly manipulate all diagram aspects	More flexible	Danger of cluttered diagrams Unwanted syntactic constructs possible

Table 3.2: Comparison of layout control approaches in graph style visualization

Another important aspect of interaction is the management of large argument maps. Over time, an argument can grow quite a bit, thus making it difficult for students to follow and review. For instance, Karacapilidis & Papadias (2001) reported that in the Hermes system, most of the discussions comprised 20 to 40 contributions. Because graph representations typically use a lot of screen space for their nodes and links, it is important for an argumentation system to offer interaction techniques that assist users in maintaining the overview of a larger ongoing argumentation, while at the same time allowing users to focus on specific (and smaller) parts of the argument. The most common techniques for summarizing large, graph-style maps are zooming and scrolling (e.g., ArguNet (Schneider et al., 2007); Rationale, and Athena). However, dependent on screen size, users of zooming and/or scrolling displays are unable to read or edit the content of an argument map if, for instance, the zoom-in level is too low. On the other hand, scrolling through large maps can cause the user to lose the "big picture".

One of the more sophisticated methods is a mini-overview map; a smaller version of the entire argument, typically put in one corner of the screen. One example of a system that uses this interaction technique is Rationale. This approach has the advantage that it can be used in combination with other techniques such as zooming. However, while it is helpful for getting a map overview, a mini-overview map does not differentiate between important and less important argument parts, which users may want to treat differently.

Filtering — that is fading selected parts of an argument based on a specific rule is another interaction technique for managing large arguments. It is used, for instance, in Athena, Convince Me, and Digalo. In Athena, it works as follows: After building an argumentation map, the user can assign scores regarding the acceptability of a statement and relevance of a relation between two elements. In Figure 3.3, the filling level of nodes (that is, the degree to which they are colored) shows the acceptability of statements, and the relevance level for relations is shown through the percentages on the edges. Based on these values, Athena calculates an overall score per element by aggregating the acceptability values of child elements, weighted according to their relevance. Elements with scores below a user-defined threshold are filtered out. For example, Figure 3.3 shows a map without filtering (0%) as compared to the same map with a 30% filtering. Especially for large graphs, such a mechanism can be useful in helping users focus on the main points of an argument. Other systems, such as Digalo, allow filtering by other criteria such as timestamps, contribution types, or author of nodes.

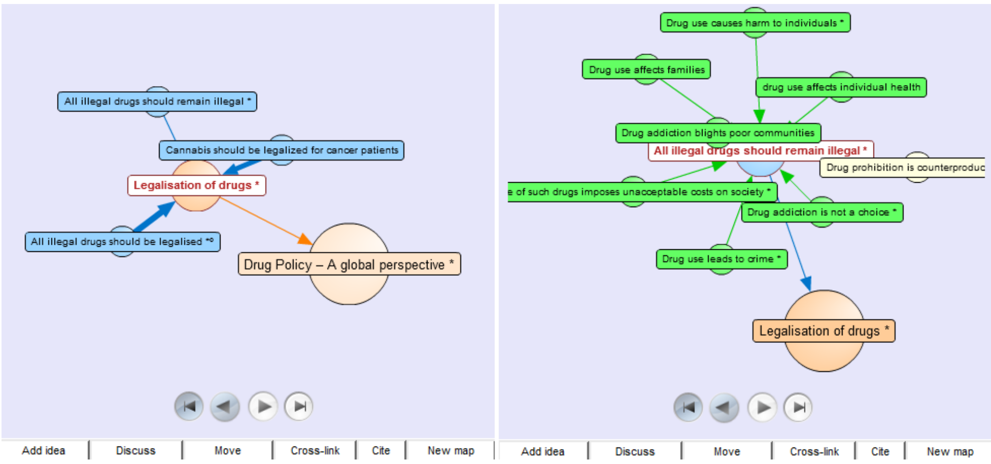


Figure 3.4: Local views in Debategraph

Another technique for reducing the visual complexity of large arguments is the use of local views that hide portions of arguments based on a "distance" to an explicitly set focus point. A prominent example that employs this technique is Debategraph. In this system, which makes use of a tree-based visualization, users see only the child nodes of their current main focus. The child nodes can also be used for navigation: When a user clicks on a node, he or she will see one level deeper in the tree, plus the parent node for reverse navigation (see Figure 3.4, which shows navigation to the "*All* *illegal drugs should remain illegal*" sub-argument). A similar interaction technique is frequently used in structured discussions and forums, in which users can expand and collapse different threads.

3.4 Collaboration

One of the critical differences between systems is whether collaboration is supported and, if it is, what types of student-to-student interactions are supported. In particular, while single-user argumentation tools support knowledge construction and learning of argumentation constructs, collaborative argumentation systems typically provide learners with insights into the opinions, viewpoints, and knowledge of other students. The range of collaboration options that are available are:

1. Single-user argumentation systems are software modeling tools that help individuals structure their thoughts and/or prepare argument representations. Some systems provide only modeling facilities, while others actively provide feedback (Examples with feedback: Convince Me, LARGO; examples without feedback: Athena, Araucaria, Carneades)
2. Small group argumentation systems serve as a software mediator between a relatively small number of learners (typically 2 to 5) and offer (typically) synchronous communication and/or collaborative modeling tools. Users may profit from both interaction with the system and with other users, developing general argumentation skills, discussing different points of view, and/or learning skills of persuasion (Nussbaum et al., 2007; Keefer et al., 2000). System-generated feedback can support both argument aspects and communication aspects. (Examples: Digalo, QuestMap, Belvedere, AcademicTalk)
3. Community argumentation systems are, in many respects, similar to small group argumentation systems but with support for a larger number of participants (and, typically, a larger number of contributions). The larger number of users puts additional constraints on the system: Communication is typically asynchronous to avoid coordination problems. The representational format enforces a more rigorous organization of contributions, that is, discussion/argument threads are preferable over graphs. (Examples: Debategraph, Collaboratorium, Debatepedia)

Of course, argumentation systems that focus on scaffolding individual work can also be used by small groups that share a single computer (cf. van Gelder (2003)), and multi-user systems can also be used by single users (e.g., Digalo). Single-user and collaborative activities are sometimes part of the same activity sequence: In some phases, individual work may be beneficial (e.g., the initial representation of one's point of view), while in others, collaboration with other students may be more fruitful (e.g., discussion of different points of view). In section 3.9, I will look deeper into the issue of phased activity sequences and their impact on learning and system use. When users try to collaborate in synchronous mode while spatially separated, they might run into problems due to a lack of coordination. One way to address this problem is to use floor control to guide turn taking. For instance, in one configured use of Digalo (Schwarz & Glassner, 2007), students must explicitly request control over the shared workspace, then perform their actions (e.g., add a new contribution), and finally release control to let the next student take a turn (a study by Schwarz and Glassner concerning the effects of floor control is discussed in section 3.9). Some systems allow "indirect" user-to-user interactions. In LARGO, for instance, student-to-student interactions take place only by students rating contributions of their peers; in Araucaria, users can share the diagrams they produced when analyzing a transcript via a central server.

Nevertheless, effective collaboration does not occur by nature. Therefore, the argumentation process can be guided by additional means which will be described in the next section.

3.5 Process Design

In sections 3.2 and 3.3, the impact of different argument representations and interaction designs has been emphasized. Nonetheless, it is important to see the context to which the representations are applied and when the interactions take place. A typical set of questions that need to be answered here are: What is the goal of the argumentation? Who participates at the discussion and what is he or she intended to do? What is the role of the argumentation tool and its representation? — Overall, the argumentation process needs to be structured.

An approach to actually structure learning processes in CSCL in general is the use of collaboration scripts (Fischer et al., 2006). In the literature, they are defined as following:

> *"[...] a **collaboration script** can [...] be described as an instructional means that provides collaborators with instructions for task-related interactions, that can be represented in different ways, and that can be directed at specific learning objectives. These objectives can be reached by inducing different kinds and sequences of activities, which are implicitly or explicitly clustered to collaboration roles. Scripted activities can be broken down into individual acts that together form a larger activity, and scripts can vary with respect to how much structure they provide." (Kollar et al., 2006, pp. 162 f.)*

Based on this definition, Kollar et al. (2006) described the following five central conceptual components: (1) *learning objectives*, (2) *activities*, (3) *sequences*, (4) *roles*, (5) *representations*. Here, I want to transfer this general concept to argumentation. To exemplify these components, Table 3.3 shows parts that could be combined to define a script in the area of legal argumentation learning. In the following, the components will be described in more detail.

Component	Script
Learning objectives	Learn how a trail works, learn to state hypotheses, learn to provide evidence for your hypotheses, learn to refute hypotheses
Activities	Swearing-in, gathering of evidence, evaluation of evidences, pleadings
Sequences	Collect arguments for the plaintiff - Collect arguments against the plaintiff; Let the parties present their sides - Judge based on the presented evidences
Roles	Judge, plaintiff, defendant, witness
Representations	Questions (e.g., "Can you refute the statements of the other party?"), Use of argument patterns (e.g., "Hypothesis is followed by facts")

Table 3.3: Possible parts of a scripted approach in legal argumentation

Learning objectives The objectives of a script in the field of computer-based argumentation can be twofold (according to section 2.2). On the one hand, there is the arguing to learn perspective (see chapter 2.2.2), that is the objective is to gain knowledge in the area in which the argumentation takes place, for instance physics, chemistry or law. On the other hand, there is the learning to argue perspective (see chapter 2.2.1), that is learners should learn how to argue effectively. As part of the latter approach, the goals are manifold, including the correct application of argumentation pattern such as hypothesis-fact-conclusion as well as dealing with problems that are typical for group work (see chapter 4.2.4).

Activities In order to achieve the objectives, there is a set of activities especially designed for this purpose. The activities define the learning phases as well as the tasks in a concrete phase. The degree of freedom of how to perform a specific activity may vary (Kollar et al., 2006). In argumentation, typical activities are the stating of hypotheses, the provision of facts as well as the drawing of conclusions.

Sequences Whereas the activities define *what* to do, sequencing describes *when* and *in which order* it should be done. For instance, the modeling of arguments can be done before the discussion individually or together during the discussion (Munneke et al., 2003). Another example of sequencing in argumentation can be found in (Schwarz & Glassner, 2007) who evaluated the use of controlled turn-taking (*floor-control*) in synchronous collaboration settings (see section 3.9).

Roles By assigning roles to group members, the participation of individuals can be enforced. Even though the overall effects of roles are unclear (Webb & Palincsar, 1996), recent research (Strijbos et al., 2004) indicated that the use of roles increases students' awareness of collaboration as well as their awareness of their group's efficiency. Compared to groups that do not use role assignments, the amount of task-focused statements is higher. In addition, the distribution and rotation of roles among group members, offers students the possibility to consider problems from multiple perspectives. In the area of argumentation, an imaginable approach here is switching between parties that are arguing pro or against a concrete position. By means of this approach, students are forced to deal with arguments of both sides. Especially in ill-defined domains such as argumentation (see section 2.3), it is important to consider multiple perspectives (Spiro et al., 1991) to find acceptable solutions for argumentative problems.

In argumentation, typical roles may include *starter*, *moderator*, *theoretician*, *summariser*, and *source searcher* (De Wever et al., 2010; Schellens et al., 2007) or *composer*, *elaborator*, and *integrator* (Nussbaum et al., 2007).

Representations Finally, an adequate representation must be chosen to communicate the script's instructions to the learners. This may occur by means of textual instructions (e.g., in the form of questions such as "Can you summarize the most important points of the argument?") or as part of the ontology (see section 3.6), which enforces argumentation patterns such as *claim*, *grounds*, and *qualifications* (Stegmann et al., 2007) or schemes (Walton et al., 2008). Another approach that has been applied in practice is the use of *argumentation vee diagrams* (Nussbaum et al., 2007; Nussbaum & Edwards, 2011). Here, the pro and con side of an issue are presented side-by-side and the learner is encouraged to ponder between both sides, before drawing a conclusion. By means of additional questions, the pondering is aided. The basic structure of an argumentation vee diagram is shown in Figure 3.5.

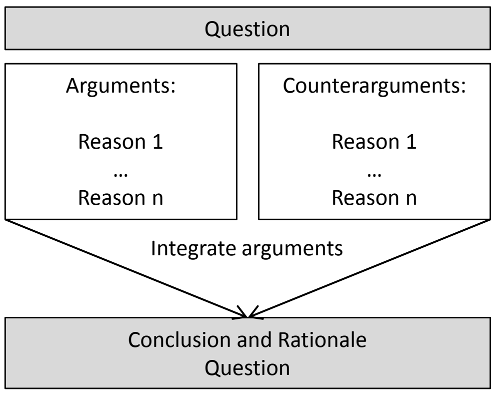


Figure 3.5: Basic structure of an argumentation vee diagram based on Nussbaum et al. (2007)

However, even though the differentiation by Kollar et al. (2006) is generally applicable, that is, to face-to-face learning as well as to computer-mediated learning, there are additional established differentiations that use different factors to categorize scripts. One of them is the differentiation presented in (Dillenbourg & Hong, 2008). Here, scripts are categorized by means of their granularity into micro- and macro-scripts:

> *"**Micro-scripts** are dialog models, mostly argumentation models, which are embedded in the environment [...] and which students are expected to adopt and progressively internalize.*
>
> ***Macro-scripts** are pedagogical models, that is they model a sequence of activities [...] to be performed by groups." (Dillenbourg & Hong, 2008, p. 7)*

Whereas macro-scripts will usually make use of most or even all components that are described by Kollar et al. (2006), micro-scripts do not. In fact, most micro-scripts consist only of the definition of a learning objectives and representations that is typically a small learning unit that should be internalized and a representation that is suitable to achieve this goal. Thus, multiple micro-scripts can be combined in activities and sequences of a macro-script that has a wider scope.

Another differentiation proposed by Schellens et al. (2007) can be made by the script's objective (content or communication) that is intended to be scaffolded:

> *"The **content-oriented scripts** are about provoking or stimulating specific cognitive activities [...] they have an immediate effect on the ongoing cognitive processes.*
>
> *The **communication-oriented scripts** aspire to stimulate the collaboration processes. They aim at fulfilling the necessary conditions for effective collaborative learning within a virtual learning environment." (Schellens et al., 2007, p. 228)*

Other examples of (collaboration) scripts are the *Jigsaw*[5] (Aronson, 1978; Aronson & Patnos, 2011) method as a general approach that is used as basis for many scripts (Kobbe et al., 2007) or more specifically for argumentation, the *ArgueGraph* script (Jermann et al., 1999; Dillenbourg & Jermann, 2006). In the former, that is the Jigsaw method, the lesson is divided into segments. Each group member is assigned to one of these segments and has no access to the other segments. By presenting one's segment to the others, each group member has an important role to fulfill. In the latter, that is the ArgueGraph script, the goal is to answer a questionnaire. First, students have to answer it on their own. Their results are presented in a two-dimensional graph. After that, pairs of students with conflicting opinions are formed. They have to fill out the questionnaire again, whereas both are aware of their prior answers. Thus, an evaluation of both positions is required. The common answer is then again presented in the graph to show the migration of each student's position. Based in this information, the teacher is finally debriefing the students.

However, the use of scripts should be carefully evaluated, because it may be detrimental in certain situations: Each script causes additional cognitive loads on the learner's side. In fact, argumentation is not a trivial endeavor at all and it is getting worse when collaboration is required. As a consequence, learners may be overwhelmed by the challenges they are faced with. Thus, an *overscripting* (Dillenbourg, 2002) must

[5] http://www.jigsaw.org/

be avoided. A more in-depth discussion of scripts (not restricted to argumentation) can be found in (Kobbe et al., 2007; Kollar et al., 2006).

3.6 Ontology

Ontologies — explicit specifications of a conceptualization (Gruber, 1993) — provide the foundation for most argumentation systems. Such representational systems describe the components of arguments, together with relations between components and modifiers of the components and relations (such as scores or relation relevance values, as shown as percentages on the links in Figure 3.3). While ontologies have most typically been used for computational reasons in artificial intelligence (AI) systems, for instance, as tools for automated logical inference, argumentation systems also use ontologies, with the aim of making users aware of the conceptual components in the task domain (Suthers, 2003).

Ontologies may combine theoretical perspectives (e.g., Toulmin (2003)) with pragmatic considerations (e.g., understandability by a specific target group). For instance, Stegmann et al. (2007) simplified the original Toulmin model, perhaps the most well-known and popular argument ontology, to improve the usability of their argumentation interface. Suthers (2003) changed the "perspective" of Belvedere to a simpler version of evidential reasoning; this was done based on the observation that students had problems using the more expressive ontology. Schwarz & Glassner (2007) say that such ontologies that result from an evolutionary and reflected process can be classified as educated ontologies; they are learned in schools and universities in the form of definitions and rules. This contrasts with informal ontologies, which are based on reasoning that typically occurs in natural conversations. While educated ontologies seem especially appropriate for argument modeling, their informal counterpart may be more suited to support structured- and typical less formal-communication. One variation is sentence-opener interfaces, which do not explicitly expose categories but which scaffold new contributions through predefined sentence-starting phrases. Typically, these interfaces are based on an underlying model of desired communication acts and processes, for instance, dialogue games (McAlister et al., 2004). One general problem that communication ontologies and sentence openers strive to address is to help students stay on topic by limiting user options.

There are a number of argumentation systems that aim at supporting a broad range of various argumentation domains with a "one-ontology-fits-all" approach. Interestingly, the ontological approaches between these systems differ considerably. One extreme case is Rationale, which employs a fine-grain ontology with approximately 30 different contribution and relationship types. While such a general approach has appeal, some studies show that such wide-ranging ontologies may confuse users with a "*plethora of choices*" (Suthers, 2003, p. 34). Other researchers have found that users of argumentation software tools are able to use effectively fairly large ontologies - of more than, say, ten elements — if 'labeling' of the contributions is done through titles of contributions, and not by selecting specific contribution types from a palette (Jeong, 2003). The other extreme is coarse-grain ontologies with unclassified nodes and a limited number of link types (e.g., pro and con), as in Athena. Such an approach is, of course, quite general, but also limited in expressiveness. A compromise is a basic set of elements that is extensible, as in, for instance, Aquanet (Marshall et al., 1991). This approach has the advantage of flexibility without overwhelming users. Yet, while quite flexible, this approach may make automated analysis techniques more difficult to implement. Table 3.4 compares the ontologies of Athena and Rationale with respect to their contribution and relationship types.

System	**Contribution types**	**Relationship types**
Athena	Node	*Unclassified*: Link
Rationale[a]	*Basic Boxes*: basis, assertion, by definition, case study, common belief, data, event, example, expert opinion, law, media, personal experience, publication, quote, statistic, web *Grouping*: box *Reasoning*: contention, reason, objection *Advanced Reasoning*: contention, reason, objection, co-premise *Teacher tools*: feedback notes (general, claim, structure, evidence, evaluation, blank) *Extras*: note, topic, question, option, pro, con, idea, consequence, information required	General links that are classified automatically depending on the types of nodes they connect

[a]Rationale organizes contribution types in its user interface according to a common theme. The table shows these themes in *italic* letters.

Table 3.4: Different approaches to achieve generality via ontologies

In general, the absolute number of ontological elements is not the only important factor. The elements' understandability, their differentiability (that is, is the difference in meaning between elements clear?), their organization in the user interface (that is, are they organized in a clear and consistent way, for instance, by grouping related elements together?), and synchronicity of interaction (that is, how much time does the user have to choose?) are also important. The development of a suitable ontology is a critical aspect in the design of an argumentation system and might involve iterative refinement based on observed problems and weaknesses (Suthers et al. (2001); Buckingham Shum et al. (2002)).

While the ontologies discussed above are intended to be quite generic, there are also argumentation tools that provide domain-specific ontologies, that is, representations that provide support for the particular requirements of a specific domain. For instance, argumentation in the legal domain is quite structured and specific, with well-defined types, roles and "outcomes" (the decision of the case). These domain characteristics are then also reflected in some argumentation ontologies (e.g., Wigmore (1931) diagrams in Araucaria). Table 3.5 shows the domain-specific ontologies of the systems LARGO, Convince Me, and AVERs (Bex et al., 2007).

System	Domain	Contribution types	Relationship types
LARGO	Law/hypothetical reasoning	Test, hypothetical, fact	Model-specific relations: Test modification, distinguish hypothetical, hypothetical leading to test change
Convince Me	Science	Hypothesis, Evidence	Explanation, contradiction
AVERs	Crime investigation	Claims about legal cases (Pro/Con), quotes from source documents, inferences, schemes	Directed links

Table 3.5: Domain-dependent elements of ontologies

Differences between ontologies in argumentation systems derive not only from domain differences, but also from different conceptualizations and perspectives of system designers. Even systems targeted at the same domain can, and do, have different ontologies. For instance, although Carneades and LARGO are used to model legal arguments, they have different ontologies: Carneades models arguments as Premise-Conclusion trees with proof standards assigned to propositions, while LARGO takes the perspective of hypothetical reasoning. Thus, ontologies are not unique even within a single domain.

Independent of an ontology's granularity and generality, there are different property types used in ontologies. Together, these properties define the aspects of an argument that are represented in a system. The properties can be divided into the following categories: (1) Content data, which contains basic argument data such as title, label, and argument text; (2) Evaluation data, numeric or categorical values to assess the believability or relevance of an element. Such data can be used, for instance, for automated filtering of content. Other types of evaluation data are, for example, the assignment of proof standards (Gordon & Karacapilidis, 1997) as used in Carneades or argumentation schemes (Walton et al., 2008), as used in Araucaria, to mention just a few; (3) Awareness/collaboration information, which is important in multi-user systems to assure that each participant is able to see, for instance, who has added each contribution; (4) Technical meta-data, such as unique identifiers for argument elements.

3.7 Supporting Learning Activities via Artificial Intelligence

To promote learning, it is essential to provide guidance for the learners. To guide learners, basically two steps are required: (1) Analysis of the modeled argument, and (2) provision of feedback.

3.7.1 Analysis of the Modeled Argument

In the first step, the argument created by the learner needs to be evaluated. However, as mentioned in section 2.3, argumentation is ill-defined. Thus, analyzing the modeled argument is hard and most classic Intelligent Tutoring Systems (ITS) are not capable of dealing with it without help of the system's users.

Overall, an analysis of the quality of the modeled argument can be done on two levels:

1. The content level
2. The structural level

The first approach is only hardly doable and error-prone. Even though, an AI software named Watson[6] recently beat two humans in Jeopardy!, a game show that requires the understanding of questions, the challenge is still open. A glance at the specifications of the machine running Watson (a cluster of 90 servers with a total of 2880 processor cores and 16 terabytes of RAM[7]) reveals that the everyday use of such an approach is neither feasible today nor in the near future.

The second approach, conducting analyses and providing feedback based on the structure of the modeled argument, is more suitable. Here, one can use the information provided during the argument modeling. In the graph-based visualization for instance, the arguments are structured as nodes. These arguments are in relation to each other by means of vertices. The nodes and vertices are categorized by means of the ontology, that is nodes may be hypotheses or facts and relations may be pro or con. In combination with additional meta-information (e.g., user-assigned scores; see section 3.6) an artificial intelligence engine is able to find possible weaknesses of the argumentation.

An example of a structural analysis approach can be found in LARGO (Pinkwart et al., 2006, 2007, 2008). Here, graph grammars have been applied to analyse the structure of the modeled argument with respect to weaknesses such as a hypothetical node that is not proven or falsified by facts. Another attempt to analyze the argument structure is done in the ECHO approach of Convince Me (Schank, 1995; Schank & Ranney, 1995). Here, the students assign scores to parts of the argument graph. These are then used to compute the internal consistency of the argumentation. Evaluations of both methods will be presented in section 3.9.

However, even though the content-based analysis is not feasible today, there have been attempts to derive the actual meaning of an argument part by means of the structure. An example here is, again, LARGO. The actual task in LARGO is to extract the argument from a given text (the transcript). To do so, the learners may relate argument parts modeled as nodes in the graph-based visualization to the given text. By means of this relation, the system is able to identify multiple solutions from different students for the same passage of the transcript. These solutions are then

[6] Watson is a part of the IBM DeepQA project. Further information about the project can be found at: http://www.research.ibm.com/deepqa/

[7] http://www.cs.umbc.edu/2011/02/is-watson-the-smartest-machine-on-earth/

presented for assessment to other learners, who have been working on the same part. Based on these assessments, a quality heuristic is calculated through collaborative filtering to actually judge the quality of the argument.

3.7.2 Provision of Feedback

Once the method to analyze an argument is decided, the results must be presented to the learner. Here, a couple of questions must be considered:

To Whom Should Feedback Be Provided?

As a first step, it must be clear to whom feedback should be provided, because this will directly influence the timing of feedback as well as the kind of feedback representation (see below). Typical recipients are, on the one hand, the students who actually argue by means of the tool (e.g., in LARGO and Convince Me) or, on the other hand, a moderator who usually is able to filter the feedback and intervene only if the feedback is actually appropriate, e.g., in ARGUNAUT (Asterhan & Schwarz, 2011). Depending on the recipient, the purpose of the feedback differs: Whereas *direct feedback* to students is usually aimed to scaffold the learning processes, *indirect feedback* to moderators may provide assisting functions to allow a human user to keep track of multiple discussions at once.

However, in combination with a possible role distribution among arguers (see section 3.5), the number of possible recipients may grow (depending on the number of roles) and the purpose of the feedback may vary.

When Should Feedback Be Provided?

Once, the recipient is clear, there are basically four timing strategies for feedback: (1) immediate or permanent feedback, (2) on demand feedback, (3) moderator-driven feedback, and (4) conclusive feedback. An overview of the pros and cons of the different timing strategies are summarized in Table 3.6.

The first timing strategy (immediate or permanent feedback) is the favored method for indirect feedback that is used to assist a moderator or teacher in supervising multiple discussions at once. Even though there might be a high count of irrelevant or misleading feedbacks (e.g., if a student states hypotheses first and plans to add facts later — here, the feedback may highlight the missing facts), a moderator usually has additional context information that allows him or her to filter those false feedbacks out. If this timing is used for direct feedback, it can help to scaffold the *current* activity (Shute, 2008). However, there is a risk of hindering students from reasonable argumentation due to excessive amounts of feedbacks that may distract students from their actual work. In addition, too frequent feedback may lead to ignorance from the student's

When?	Advantages	Disadvantages
Immediate / Permanent	All detectable mistakes are revealed	May distract users from their work
	Beneficial to improve current activity (Shute, 2008)	Feedback may become excessive
	Content is still present	Not always possible, because the context may dependent on other, not yet stated parts
On demand	Does not distract users from their work	Important feedbacks can be missed
	Frequency is user-dependend, flooding avoided	
	Less authoritative, preventing discouragement (Suthers et al., 2001)	
Moderator-driven	False and irritating feedback can be filtered	Limited feasible for a large set of users
	Frequency can be adjusted on-the-fly	
Conclusive	Does not distract users from their work	Error context may be missing

Table 3.6: Feedback timing strategies

side to keep the cognitive load manageable. This is especially true if the feedback is obvious or wrong in some cases. Examples of systems that make use of this approach are ARGUNAUT, Group Leader Tutor (Israel & Aiken, 2007), and Pierce (Goodman et al., 2005a).

The second timing strategy (on demand feedback), follows a different, more active model. Here, the user decides when he or she wants to get feedback. According to Suthers et al. (2001), this approach prevents discouragement of students, because it is less authoritarian compared to other timing methods. Obviously, a continuous flooding with feedback messages will be prevented. However, there are potential disadvantages as well: Important feedbacks that may be helpful in certain situations may be missed. Since students are typically not aware of their own mistakes, they may not ask frequently enough even when they are stuck, as shown in the re-evaluation of the LARGO system (Pinkwart et al., 2008). In addition to LARGO, Belvedere is another example of a system that follows this timing approach.

The third timing strategy (moderator-driven feedback) is a hybrid approach between the first and second one. Whereas the moderator will usually get immediate or permanent feedback (timing approach 1), he or she decides when the feedback is sent to the student. The active part of the second approach is, hence, re-located to the moderator's side. Thus, this approach benefits from the advantages of the other approaches. On the con side, it is only doable for a limited set of users and discussions and requires additional personal that is experienced with moderation of argument discussions. An example system that use this approach is ARGUNAUT (McLaren et al., 2010; de Groot et al., 2007)

The fourth and final timing strategy (conclusive feedback) divides the argumentation process more strictly. Once the argument modeling task is completed, there is feedback provided, but not in between. This approach is similar to classic exercises in class: A student gets a task, solves it and waits for the assessment of his or her solution. Depending on the concrete application of this timing strategy, the feedback may be considered in a next iterative step. The clear advantage here is that the student does not get distracted from his or her work and may have time to reflect about the feedback afterwards. However, the context in which the error occurred as well as the steps that may have led to the mistake can be lost or has to be reconstructed, which makes it hard for the student to learn from it. Example systems that make use of an (iterative) approach with conclusive feedback are Convince Me and Group Leader Tutor (Israel & Aiken, 2007).

How Should Feedback Be Provided?

Finally, an important aspect of the AI support is the way how the computed feedback will be reported to the recipient including the following:

Prioritization The number of possible feedbacks can be large. However, to avoid a cognitive overload of the recipient, it is important to filter the feedback. One way to do so is the prioritization of the computed feedbacks. Examples can be found in Belvedere (Suthers et al., 2001) and LARGO (Pinkwart et al., 2006). The former uses a preference-based quick-sort algorithm, which iterates through a list of criteria which includes, e.g., the priority of new advice and priority of expert advice over syntactic advice. The latter assigns argumentation phases to each analysis pattern. Feedback that is connected to the phase in which it is requested will be preferred. In addition, the current view — for instance the visible part of the transcript — will be considered. Thus, feedback that is assumed to be appropriate in this phase will get a higher priority.

Visualization Once the appropriate feedback has been selected, the way how to present it to the recipient is the next thing to decide. Existing approaches include the following: (1) textual representations, (2) the highlighting of argument parts as well as (3) the use of visual displays.

The first approach (textual representation) is the most common way to present feedback (Suthers et al., 2001). Here, text messages will be shown to the users that explain the identified problem and give hints how to solve them. In these textual messages, concrete connections to elements of the modeled argument can be integrated (e.g., “fact number 6 is not connected to any hypothesis”).

Directly connected to this is the second approach (highlighting of argument parts). By means of highlights, the elements that the feedback aims for can be made salient to the user. However, even without additional textual feedback, highlights can be used — for instance by a moderator (as done in ARGUNAUT (McLaren et al., 2010)) — to draw the students' attention to a specific part of the argumentation.

The third approach (meters) are used for instance in ARGUNAUT (McLaren et al., 2010). By means of meters high-level analysis of the argumentation parts (e.g., indicating the number of elements per user) can be made visible to a moderator, who is then able to draw conclusions (e.g., a student is not participating at the discussion) and provide additional guidance to students. Further, the results of students' interactions can be *mirrored* to increase students' awareness about their actions and behaviors (Soller et al., 2005).

Intention & Formulation As mentioned before, the provision of feedback is hard and, hence, error-prone. Thus, it is important to make sure that a possibly wrong feedback may not confuse the recipient or cause uncertainty, which could happen if the feedback is formulated to strict, e.g. in form of a concrete instruction (for instance, "The hypothesis is not supported by facts, delete it."). To avoid this, feedback should be carefully formulated, e.g., in form of hints (Suthers et al., 2001) or suggestions (Pinkwart et al., 2006). The basic intention behind this is that the feedback should encourage the student to think about the modeled argument and possible weaknesses of it (Suthers et al., 2001). To go one step further, in LARGO self-explanation prompts (that have shown to be effective, e.g., in (Schworm & Renkl, 2007)) have been used to promote the self-reflection of students.

However, the topic of analysis and feedback in the domain of argumentation is a research area on its own and cannot be covered with all its facets here. A more detailed review can be found in (Scheuer et al., 2010, 2011).

3.8 Architecture & Technology

Another key aspect of argumentation systems is their underlying software architecture. Building upon a solid software architecture is beneficial in reducing development time and in achieving stability, extensibility, and performance. Even more importantly, a suitable software foundation is critical to implementing technologically sophisticated CSCL settings which use collaboration scripts (Weinberger et al., 2005; Kobbe et al., 2007) (see section 3.5) and/or floor control, both of which have been shown to be promising for learning in the context of argumentation (see section 3.9). However, most argumentation systems are based completely on their own, unique design and code, without even reuse of past successful designs. The review, which included e-mail contacts with the developers of 12 systems asking them about system architectures, also revealed a huge lack of software documentation. As such, reusable design patterns and guidelines have thus far not emerged from the field of argumentation systems. In comparison, in the more general field of software engineering, design patterns have been commonly used in system development for about 15 years (Gamma et al., 1995).

The few publications about software architectures for educational technology systems include early work by Wenger (1987), who proposed an Intelligent Tutoring System (ITS) architecture based on four software modules (expert, student, tutor, and communication model). This approach recognizes the significant advantages of clearly separating functionality and designing systems based on a well-specified and modular architecture. However, this early and general proposal, while followed since then in the field of ITS, is not enough for modern, distributed, and collaborative educational technology systems. The main contributions of more recent work in the field will be summarized in the following.

Suthers (2001) discussed different generic argumentation architectures, based on experiences from different versions of Belvedere. He differentiated systems based on their coupling model: (1) strict "What you see is what I see" (WYSIWIS), (2) relaxed WYSIWIS, where different users can have different viewports on a shared view, and (3) model-level coupling in which users see the same semantic state of a shared model, but the views may be totally different. Comparing a centralized architecture (used in Belvedere v1) and a mixed replicated/distributed architecture (Belvedere v2), Suthers finally proposed a hybrid architecture that combines the advantages of the different architectures: The model is available on a central server as well as, in form of a copy, on the client machines. Furthermore, it is possible to have different views on the same underlying data. Therefore, users are able to choose the view that best fits their needs at any time without losing the possibility for collaboration with others that use a different view.

Harrer & Devedzic (2002) and Devedzic & Harrer (2005) have identified some design patterns for ITS systems, based on detailed reviews of existing systems. Examples are the KnowledgeModel-View pattern that manages multiple models and views (analog to the MVC pattern for one model and view) and the ICSCL pattern, which allows adapting learning materials separately for individuals and groups at the same time. Even though these patterns are described in the context of ITS, some (including the two mentioned above) can be applied to designing collaborative argumentation systems (e.g., to give user-dependent feedback).

While Harrer and Devedzic report on general design patterns, some other publications describe the design of specific systems. Goodman et al. (2005a) show how Pierce, an agent for supporting collaborative learning, is designed, and Israel & Aiken (2007) present the architecture of their "Intelligent Collaborative Support System". Bouyias et al. (2008) report on ideas for an architecture to support the fading of collaboration scripts. Also, some authors propose component-based architectures for their ITS systems (Kumar et al., 2007; Israel & Aiken, 2007; Tedesco, 2003) to facilitate the exchange of modules, argue for specific client-server architectures underlying their implementations (Baghaei et al., 2007; Tedesco, 2003; Vizcaino et al., 2000), or describe architectures comprising collaborative learning tools (Belgiorno et al., 2008). Yet, these software architecture descriptions are not adapted to the specific requirements of educational argumentation systems, and it is not clear how they can be used more generally.

In addition to architecture, another important technological aspect is the format used to save and exchange argumentation data in different systems. The choice of the data format is important: A standardized and agreed-upon format, for instance, would facilitate conducting meta-analyses of study data even if the studies were done with different tools. Also, common formats would allow for interoperability between applications, enabling a data exchange (e.g., in order to load data gathered with one tool into another for analysis purposes). In addition, different formats have different affordances. Two primary approaches for argumentation data formats have thus far emerged: state-based (e.g., GraphXML: Herman & Marshall (2000) Graph Exchange Format (GXL): Taentzer (2001); Argument Interchange Format (AIF): Chesñevar et al. (2006); GraphML: Brandes et al. (2002)) and action-based (e.g., the Common Format (CF), which was used in the ARGUNAUT project). While the former approach only saves the current state of an argument, the latter stores every action, such as adding, removing, or editing parts of the graphical argument. The action-based approach uses less bandwidth (because only small updates are sent, not the whole map), and is more intuitive for collaborative systems where actions of users must be broadcasted to other users. The action-based approach, however, requires more time to compute a map state at a given time, which is required whenever a new client joins an argumentation, because all actions from the beginning to the given time have to be provided to the new client. The choice of format also holds implications on the options for automated argument analysis and feedback: Some analyzers use actions as inputs and would, thus, benefit from an action-based data format (e.g., ARGUNAUT), while others are based on the state of an argument (e.g., LARGO) and, thus, work better with a state-based data format.

In summary, even though some proposals for data formats have been made, none has been established as a standard for argumentation systems yet. One reason for this may be that they are all limited in different ways: While the mentioned graph-based formats can be used for graph-based representations, they do not work as well for other representations. Furthermore, they do not provide support for argumentation-specific needs (e.g., transcripts or links to external resources) by default. The existing argumentation-specific formats support these needs, but are not flexible enough to support the variety of argument styles used in the different systems.

3.9 Empirical Studies

In the prior sections of this chapter, various approaches to support argumentation and argumentation learning have been presented. However, even if something can be done, it does not mean that it has to be done. Instead, it is important to be aware of the consequences that result from the use of specific aspects such as argument visualizations, ontologies, scripts and adaptive feedback. Therefore, this section summarizes empirical research that has been conducted to evaluate the usefulness of these means for argumentation and argumentation learning. The presentation of the studies will follow a pattern consisting of domain, hypotheses and research questions, method and results. The underlying hypotheses are, in most cases, directly taken from the original research papers.

A brief overview of the studies and their underlying questions is shown in Table 3.7.

Area	**Reference**	**General description**
The effect of argument mapping for learning (→ 3.9.1)	Easterday et al. (2007)	Using diagrams to teach causal reasoning on public policy problems
	Carr (2003)	Using diagrams to teach legal reasoning
	Harrell (2008)	Using argument mapping without computer support to teach philosophical argumentation
The effect of argument representations (→ 3.9.2)	Suthers & Hundhausen (2003)	Comparing the effect of different representational notations
The effect of collaboration (→ 3.9.3)	Sampson & Clark (2009)	Comparing group vs. individual performance in scientific argumentation
	Munneke et al. (2003)	Comparing the effect of constructing diagrams (a) individually before a debate versus (b) collaborative during a debate
The effect of ontologies (→ 3.9.4)	Schwarz & Glassner (2007)	Using informal ontologies and floor control in graphical e-discussions
	McAlister et al. (2004)	Using sentence openers to support critical discussions
The effect of micro-scripts (→ 3.9.5)	Stegmann et al. (2007)	Using patterns to scaffold single arguments and argumentation sequences
The effect of macro-scripts (→ 3.9.6)	Schellens et al. (2007)	Using role assignments to improve argumentation quality
	Lund et al. (2007)	Comparing the effect of instructing students to use diagrams for debating versus representing a debate
The effect of adaptive support (→ 3.9.7)	Pinkwart et al. (2007, 2008)	Using LARGO (and its feedback) to teach hypothetical reasoning
	Schank (1995)	Using Convince Me (and its feedback) to improve students' reasoning skills

Table 3.7: Overview of empirical studies

3.9.1 The Effect of Argument Mapping on Learning

One of the key features of argumentation systems is the concept of argument mapping. Here, arguments are visualized, in most cases, as graph. From a theoretical point of view, there is a set of assumed benefits. Van Gelder (2003), for instance, argued that diagrams are more readable and comprehensible than textual arguments. This is true, in particular, when considering the reduced efforts required in interpreting these diagrams. By means of, for instance, different colors, shapes and spatial arrangement of elements, the structure can be recognized more easily. Thus, strengths and weaknesses are easier to see (Buckingham Shum et al., 1997). In collaborative contexts, however, visually represented arguments force arguers to make their thinking visible (Bell, 1997), that is turning abstract arguments into feasible structures. The benefit is obvious: There is a shared artifact that everybody is aware of. However, on the downside, there may be a *"cognitive overhead"*(Conklin & Begeman, 1989, p. 211) caused by diagrams as well as a premature commitment to structure (Buckingham Shum et al., 1997). Therefore, it is important to check whether the process of creating visual representations, most likely graph-based diagrams, is helpful for learning purposes. Additional potential risks include the unnatural approach itself: While people learn in school to formulate their thoughts in prose, they are typically not used to any kind of diagram representation. Finally, the assumed pros may turn into cons once a discussion lasts longer and the diagram grows.

How to use best diagrams

Whereas these are mainly theoretical thoughts, most of them lack concrete evidence from practice. A first step to clarify the benefit of diagrams is done by Easterday et al. (2007). Here, the focus is on the learning in public policy.

Domain Public policy / philosophy.

Hypotheses / Research Questions (1) For what domains and what kind of pedagogy will diagrams help?; (2) Should learners be given pre-made diagrams or should they construct diagrams on their own?; (3) Does using / constructing a diagram have the same effect on learning as it does on performance?

Method 63 students without prior training on causal reasoning enrolled in undergraduate philosophy classes participated in a four parted experiment. The students were divided into three groups (text (n=24), diagram (n=24) and tool group (n=15)). The phases of the experiment included a pre-test, a short 15 minutes training, a performance test as well as a final learning test. Each test consisted of a policy argument as well as 10 multiple choice questions. Whereas the students in the pre-test and in the final learning test were given a policy argument represented in textual form, the training and performance tests differed between groups. In the text group, the training as well as the performance test contained instructions and arguments in textual form only. For the diagram and tool groups, the interactive web pages that were used for training contained diagrams of arguments. The last difference was the performance test. Here, the diagram group received an argument as text with a diagram, whereas the tool group was given the text and the iLogos tool to actually create a diagram on their own.

Results No significant differences on the pre-test. On the performance test, there was a (non significant) tendency that the diagram group outperformed both the text group as well as the tool group. The final learning test showed that both the diagram and the tool group significantly outperformed the text group. Here, the tool group achieved a even 5% higher test score than the diagram group. A further distinction between students who created diagrams on the learning test (n=6) and those who did not create diagrams (n=57) revealed a significantly worse result for the latter.

This study shows that the creation of diagrams — here by means of the iLogos tool — can improve learning effects. However, a limiting factor, especially when considering the learning effect, is the low number of students who actually created diagrams in the test. In addition, this short-term experiment does not consider potential effects on the long-term and it did not involve any collaborative settings, which are considered to be more complex involving additional factors that may influence the overall success of the diagramming approach.

Using argument mapping to improve critical thinking skills

A similar experiment has been conducted by Carr (2003). The main difference is the use of a different tool (QuestMap instead of iLogos) in a different domain (legal argumentation instead of public policy / philosophy).

Domain Legal argumentation.

Hypotheses / Research Questions How does the use of an argumentation system (in this case QuestMap) affect the quality of arguments generated on a practice final exam? The author hypothesized that the use of QuestMap will lead to higher quality arguments compared to traditional written arguments.

Method In order to evaluate the research question, a quasi-experimental study involving 76 second-year law students was conducted. The students were assigned to two conditions: The first condition was the treatment condition. Here, 36 of the 76 students volunteered to use the QuestMap tool. The rest of the students was assigned to the control condition, whose participants were required to solve the tasks without any additional tool. Overall, there were five problems that needed to be solved during the semester in groups of 3 to 4 students. To measure the potential differences between the conditions, the solutions were assessed by means of the Toulmin (2003) argumentation scheme with respect to the number and types of argumentation structures created. In addition, the students' performances on the practice final exam assessed by the professor of the course was used for comparison.

Results There was no pre-existing differences between groups. The comparison of the dependent variables showed that the arguments did not get more elaborated throughout the semester. Further, there was no significant difference in the overall performance on the final exam. According to a final survey among participants, students in both conditions spent a similar amount of time on tasks.

Contrary to the study of Easterday et al. (2007), this study did not show any significant benefit of using argumentation systems and argument diagramming with respect to learning. However, the studies are not directly comparable because they differ in multiple aspects such as the domain in which the evaluation took place, the tool that was used and the involvement of collaboration.

Argument mapping without software

The studies by Carr (2003) and Easterday et al. (2007) showed somewhat contradictory results. One possible reason is the use of different tools. Thus, the usefulness of argument mapping per se should be evaluated independently, that is without the use of a concrete tool. One study that actually does this is the one of Harrell (2008) who evaluated if the explicit teaching of argument mapping, even without the use of computer-based tools, is an adequate way to promote learning in the area of analyzing arguments.

Domain Philosophy.

Hypotheses / Research Questions (1) Students who explicitly learned argument mapping will improve their argument analysis skills more than students who did not; (2) all students, even those who were not explicitly taught argument mapping, will improve their argument analysis skills when using it.

Method Overall, nine lectures of introductory philosophy with a total of 269 students in semesters 1 or 2 have been evaluated. Half of them were explicitly taught how to construct argument maps, half of them were not. To measure the improvements of the students, a classic pre-/post-test design was used. In the tests, one task was to create a *"visual, graphical, schematic, or outlined representation of the argument"* (Harrell, 2008, p. 7). The concrete representation used by the students was, hence, not restricted to concrete argument maps.

Results All students, even those who were not explicitly taught argument mapping, improved significantly. Further, the students who were explicitly taught argument mapping improved significantly more than those who were not. However, not all students who were taught argument mapping did in fact use it in the post-tests and there were students who used the technique even though they were not taught explicitly. A more in-depth comparison by means of the amount of argument maps constructed showed that even though the gains of these two groups were similar, the relative improvement was significantly higher for those students who constructed many argument maps on the post-test. Further, a significant difference was found between students who constructed correct argument maps in comparison to those who constructed incorrect ones. Here, the latter scored significantly lower when considering the overall gains as well as the relative improvements.

This study shows that the use of argument mapping techniques is beneficial to learn argument analyzing. Thus, it provides evidence that the general approach to represent arguments visually — as done in computer-based argumentation tools — is beneficial for learning. However, this study has several limitations as admitted by the author. First, there were no standardized tests used to measure the performance of the students. Second, the instructor of the courses differ. Having in mind that the way of teaching argument mapping is not standardized as well, this could have major impact on the results.

Overall, the studies indicate that argument mapping is likely to be beneficial for learning purposes. However, they also make clear that the context in which the technique is applied is an important factor and largely unexplored, yet. Imaginable factors that may influence the outcomes include the domain, the collaborative setting and the used means (argumentation systems or even a pen and paper) as well as even less explored ones such as gender, age or educational background.

3.9.2 The Effect of Argument Representations

As shown in section 3.2, graph-based argument representations are not the only way to visualize arguments. Instead, there are multiple kinds of visualizations of arguments available in practice including graphs, matrices/tables, container, threaded discussions or plain text. These representations are designed to aid the argumentation. Nevertheless, each of them has a set of implication, which influence the kind of help provided by the representation. While graphs emphasize the general structure of the argument, matrices/tables focus on potential relations between structures. This *representational guidance* has been further evaluated by Suthers & Hundhausen (2003), who conducted a study comparing graph, matrix and text visualizations for arguments.

Domain Scientific argumentation.

Hypotheses / Research Questions (1) The ontological bias of the representations will affect participants' use of epistemological concepts. Here, participants are assumed to classify their ideas rather in the graph and matrix condition compared to the text condition, because the latter does not provide any categorization.

(2) Participants who construct matrices will talk more about evidential relations than participants who constructed graphs, and both of these groups will talk more about evidential relations than participants who constructed text documents. Here, the prompting character of empty cells is expected to promote

thinking about how to fill them. In the graph condition, missing relations are less salient than in the matrix condition, but more salient than in the text condition.

(3) Participants who constructed matrices will elaborate on previously represented information more than those who constructed graphs, and both of these groups will elaborate more than those who constructed text documents. Similar to the second hypothesis, the salience of matrices is assumed to be higher than the one of graphs and both are assumed to be beneficial compared to text.

(4) The process differences will lead to significant differences in learning outcomes and subsequent products of the inquiry. Caused by the higher degree of elaboration expected in the other hypotheses, there is an overall improvement assumed since more elaborated information will be better memorized.

Method A controlled lab study with paid participants (32 female, 28 male) in self-selected, same-gender pairs was conducted. The participants were students from biology, chemistry, physics, and computer science courses at the University of Hawai'i, They were under 25 years old. All of them received a 10 minute introduction to the software followed up by a 12 minutes warm-up problem. After that, they were required to solve the main problem without time limit. Next, there was an individual post-test consisting of relevant multiple-choice questions. Finally, all participants were required to complete a collaborative essay using a word processor within 30 minutes.

Overall, there were three conditions. They differed in the visual representation of arguments. One condition used a graph-based visualization, one a matrix-based visualization and the last one used a simple word processor.

Results Even though the time for the main task was not limited, the mean time that was used by the participants was similar. The differences were not significant. The first hypothesis could not be confirmed. Here, no significant differences were identified. However, the introductory demonstration of the tool in the text condition used an explicit classification of the components and, hence, is likely to have influenced the participants. With respect to the second hypothesis, the results provided evidence that participants in the matrix condition provided significantly more evidential relations (25%) than participants in the graph (12%) or text condition (9%), that is the matrix representation made potentially missing relations more salient. The third hypothesis was partly confirmed. Here, significant difference was found between both graph and text as well as matrix and text. That is, matrix and graph visualizations are superior compared to a text visualization in prompting elaboration on already represented information. However, contrary to the expectations, there was no significant difference found between matrix and graph condition. A similar picture is drawn for the fourth hypothesis. Here, the graph condition significantly outperformed the matrix condition with respect to the percentage of carryover items, that is items that were represented in the problem solving session using the tool and also included in the final essay. Nevertheless, no significant difference could be found between conditions with respect to the overall learning effects.

In conclusion, these results indicate that the use of different visualization has an guiding impact on the argumentation process and outcome. However, these influences are not always beneficial. In the matrix condition, for example, the number of relations was quantitatively higher, but this does not implicate a qualitative improvement. Here, Suthers (2003) noted that the prompting character of matrices might have triggered an overly extensive consideration of relations, when many of the relations were irrelevant. Thus, the emphasized salience might be detrimental in same cases as well. Furthermore, the combination of multiple representations for multiple purposes (e.g., a graph for argument representation and a chat for discussion or communication) as done, for instance, by Suthers et al. (2008) comes along with additional complexity. These and other limitations such as the two-dimensional nature of matrices should be kept in mind as well as potential risks when using argument representations.

However, more recent research (Suthers et al., 2008; Janssen et al., 2010) confirmed the results of Suthers & Hundhausen (2003) in large parts. In addition, the work of Janssen et al. (2010) identified higher learning benefits when using a graph-based debate tool instead of a textual one.

3.9.3 The Effect of Collaboration

An important factor in argumentation is collaboration. The reasons for that are obvious. Typical problems are highly complex and an individual is most likely not able to have all possible explanations in mind. Here, the different perspectives provided by other arguers are of high value to find answers.

General Effect of Collaboration

However, most of these assumed benefits are of theoretical nature. In fact, collaboration comes along with additional efforts concerning coordination and structuring of argumentation. In order to practically test the effect of collaboration on the outcomes of argumentation, Sampson & Clark (2009) conducted a field study in a science classroom.

Domain Scientific argumentation.

Hypotheses / Research Questions (1) Do groups of students craft better scientific arguments than individuals? (2) To what degree do group members adopt and internalize the arguments crafted by their group? (3) Do students who engage in argumentation with others demonstrate superior performance on individual mastery and transfer tasks when compared to students who engage in argumentation alone?

Method Overall, 168 high school students aged from 15 to 17 years enrolled in introductory chemistry classes participated. The participants were assigned to two conditions with 84 students each. The independent variable to distinguish the conditions was collaboration, that is half of them was assigned to the individual condition whereas the other half was assigned to the collaborative condition. A pre-test showed no significant difference between groups concerning background knowledge. The experiment consisted of four sessions. Each session lasted approximately 50 minutes. In the first session, the participants were required to complete a Thermodynamics Content Knowledge Test (Sampson & Clark, 2009). The following second session consisted of an introduction to the argument framework. The third and fourth session dealt with the concrete thermodynamical problem that should be solved. Whereas the third session dealt with the core problem, the fourth used a mastery problem. Depending on the condition, the students worked during the third session in triads or alone. In the fourth session, all participants had to work alone. In addition to the mastery problem in the fourth session, the students were also required to solve a transfer task in this last session.

Results The evaluation of the first research question revealed that the average argument quality of triads is higher than the argument quality of the lowest performing individual, but equivalent to the second highest, that is middle, performing individual. However, the results of the evaluation of the second research question showed that a substantial proportion of the students in collaborative sessions adopted the arguments created within the group. Concerning the performance on individual mastery and transfer tasks, the results showed that students who worked in a group performed significantly better (with moderate effect sizes) than the students who worked alone. This is true for both, the mastery and transfer tasks.

The take away message of the study is that involvement of collaboration per se is not a guarantee for scientific learning. In order to benefit from collaboration, it is important to see it as one factor in the complete argumentation process. As Sampson & Clark (2009) noted "*group outcomes are not simply the sum of individual abilities*" (p. 475). Instead, the "*group interaction processes clearly influence group outcomes*" (p. 475).

Using Different Collaboration Settings in Different Argument Phases

To scaffold the collaboration process, it may be beneficial for the individuals to prepare before discussion in order to promote a more elaborated discussion. Munneke et al. (2003) evaluated this question by comparing two studies which used argument diagrams in different collaborative settings (individual vs. collaborative) for different purposes (as preparation before a debate and as aid during the debate).

Domain Education.

Hypotheses / Research Questions (1) How are argumentative diagrams used during electronic argumentative interaction, in particular: (a) an individual diagram constructed *before* the debate?; (b) a collaborative diagram constructed *during* debate?

(2) To what extend do students collaboratively explore the space of debate in depth and breadth using these two kinds of diagrams?

Method Two experiments were conducted. Overall, 126 pairs of 16-17 year old students from five different schools for upper secondary education in the Netherlands voluntarily participated. For the evaluation, 10 randomly selected pairs per experiment were selected. The pairs were spatially separated, when possible. Both experimental groups used the TC3 tool and consisted of six sessions *á* 50 minutes.

In the first experiment (SCALE), students were required to construct an argument diagram individually before a debate, whereas students in the second experiment (Twins) had to construct argument diagrams during the debate collaboratively.

Results With respect to the first question, the results are ambivalent. The size of the diagrams as well as their use differed a lot. However, in both experiments students provided more arguments in favor of their standpoint instead of against it. In SCALE, the diagrams consisted of 4-17 boxes and were mainly used as information source for the following debate. In Twins, however, the diagrams consisted of 9 to 16 boxes and were used, for instance, as notebook to summarize prior chats. In addition, the construction of diagrams in Twins created input for further discussion. Further, the collaboratively created arguments were mostly chaotic unless a single student was responsible for the layout.

Concerning the second research question, students in the SCALE experiment discussed more in breadth about the topic than those in Twins. That is, they were mainly talking about their own opinions without elaborating on other standpoints. However, there was no significant difference found with respect to the discussion depth.

Similar to the study of Sampson & Clark (2009), collaboration per se did not lead to improved reasoning. However, the study of Munneke et al. (2003) highlights potential problems that are likely to occur in collaboration, e.g. chaotic diagrams. This indicates that collaboration may require additional support. An open question, however, is how this support can be provided. Imaginable mechanisms could include social factors such as role assignments (see sections 3.9.6 & 3.5) or structural guidance provided by ontologies (see section 3.6).

3.9.4 The Effect of Ontologies

In section 3.6, it was highlighted that the amount of available ontologies differ quite a lot. This is true between domains as well as within. However, an important question is to clarify what the actual role of the ontology is. Are ontologies really effective to structure the argumentation process? On the one side, their beneficial role to support artificial intelligence analyses is clear. By means of a implicit categorization, the provision of feedback can be simplified a lot. The other side which is even more important is the learners side. Here, it is essential to clarify whether the provision of an ontology helps or even hinders the learning process.

The Effect of Ontology and Floor control

A study which evaluates the role of ontology is the one by Schwarz & Glassner (2007), who applied an (informal) ontology to a graph-based argument visualization.

Domain General discussions.

Hypotheses / Research Questions Hypotheses concerning the factor *ontology*: (1) The use of an informal ontology (consisting of *claim*, *information*, *explanation*, *argument* and *else* represented by means of different shapes in the graphical environment) is beneficial since it invites discussants to be explicit about the role of their intervention in discussion. (2) The explicitness and reflection will lead to more relevant claims and arguments as well as greater reference to other students.

Hypotheses concerning the factor *floor control*: (1) Floor control, that is an enforced turn-taking mechanism to simulate asynchronous discussions, will invite discussants to delay their reactions, to plan interventions, to post them when ready, to refer to previous interventions, and to elaborate more relevant claims and arguments. (2) Synchronous e-discussions without floor control will lead to more chat-style writing, especially when no ontology is available.

Method A quasi-experimental 2×2 design with 54 grade seven students from two classes in the same school in Israel were chosen. The independent variables were ontology and floor-control. The experiment was conducted in a dialogic thinking and technology course. The students were distributed among 12 discussions, that is 3 per condition, with 3 to 6 students each and they were spatially separated. In order to measure the differences between conditions, the dependent variables included the number of (a) relevant claims, (b) relevant arguments (that is, reasoned claims), (c) chat-style expressions, (d) productive references to peers (that is, references accompanied by a new relevant claim or argument), and (e) other references to peers.

Results The amount of chat-style expressions in groups without floor control and ontology was significantly higher than in the condition with ontology or floor control. In fact, the condition who used ontology and floor control provided the fewest chat-style expressions. A similar picture is drawn concerning the number of “other references”, that is non-productive references, to peers. Here, the groups without ontology and floor control performed inferior, that is provided more "other references" than the group which were provided with an ontology and floor control.

Even though there were no overall significant effects found with respect to the factors "relevant claims" and "relevant arguments", the authors report that a LSD post hoc test showed a significant difference between the condition without ontology and floor control and the one provided with both. Again, the group supported with both scored significantly higher than the group without ontology and floor control.

This study provides evidence that the use of an (informal) ontology is indeed beneficial for the overall argumentation process. Here, the clear benefit is to keep the arguers on track, that is to avoid off-topic discussions. However, open issues are still present: Is the granularity of an ontology important? And, is there a "best" ontology for certain domains?

The Provision of Sentence Openers in AcademicTalk

While the study of Schwarz & Glassner (2007) showed that an ontology can be beneficial within a graph-based environment, it is not obviously clear that the effects can be transfered to other argument representations. In order to clarify this point, the study by McAlister et al. (2004), who used a similar approach in a threaded discussion environment, will be presented next.

Domain General discussions.

Hypotheses / Research Questions Not explicitly stated.

Method Four discussion groups consisting of a total of 22 students volunteered to participate at a quasi-experimental study as part of on online university course. Two of the groups used a chat, while two used the AcademicTalk argumentation tool. The latter provided the participants with a thread-based visualization of arguments. In addition, the participants were required to categorize their statements by means of a given set of sentence openers. In order to promote typical argumentation patters, best fitting sentence openers were highlighted. Both condition went through a four-phased activity consisting of the following phases: (1) individual preparation, (2) exploratory discussion, (3) controversial discussion, and (4) summary.

Results The AcademicTalk condition outperformed the chat condition in multiple aspects. On the one hand, there were significantly less off-topic contributions present (1% compared to 28% in the chat condition). On the other hand, the amount of desired argumentation moves including requesting/referring to evidence (13% compared to 6%), reasoned claims, rebuttals and disagreement moves raised significantly.

The results essentially confirm the findings presented by Schwarz & Glassner (2007). However, in this experiment, the reasons are not as clear as in the other study. In fact, there are at least three potentially influencing factors: (1) the kind of argument representation (thread vs. chat), (2) the presence of sentence openers, and (3) the recommendation of fitting sentence openers. Thus, these results can only serve as indicator.

In conclusion, the use of ontologies can be assumed to be beneficial. By means of an enforced categorization of argument moves, typical argumentation patterns can be fostered in the arguers minds. In addition, ontologies are likely to keep the participants on track, that is to avoid off-topic within the argument. However, still open issues include the role of ontology for specific argument activities as well as the question if there is a 'best' ontology for certain domains. For the former, one can imagine to use different ontologies for different purposes, e.g., a very easy one for early brainstorming in which categories are not that relevant and a more complex one in the actual discussion. The latter is somewhat connected to it. If there are specific activities in argumentation that can be supported best using a certain ontology, this may lead to further insights in which domains which ontology is likely to be more suitable.

3.9.5 The Effect of Micro-Scripts

As mentioned in the description of process design in section 3.5, micro-scripts usually focus on small learning units that are intended to be internalized. A typical example application area is the structural level of arguments. One way to actually provide guidance is the use of templates to enforce a specific structure. Here, Stegmann et al. (2007) evaluated the role of such templates for constructing single arguments as well as for constructing argument sequences.

Domain Educational Science.

Hypotheses / Research Questions How does the presence of a script for the construction of single arguments, a script for the construction of argumentation sequences, and their combination affect (1) the formal quality of argumentation in discussion and (2) individual knowledge acquisition?

Method 120 students of Educational Science at the University of Munich randomly divided into triads participated in the study. The study followed a 2 × 2-factorial design, that is with versus without script for the construction of single arguments and with versus without script for the construction of argumentation sequences. As argument visualization threaded discussions have been used. The script for the construction of single arguments followed a simplified version of Toulmin's argumentation scheme (Toulmin, 2003). Here three text fields were given as template: claim, grounds and qualifications. However, the students were not required to use them. Instead, they were also allowed to use a free-form text field in order to provide their arguments. The script for the construction of argumentation sequences followed the proposal of Leitão (2000). Here, a typical sequence is a repetition of a argument-counterargument-integration chain. A classic pre-/post-test design was used. These tests included questions to evaluate domain-specific as well as argumentation knowledge and lasted 15 and 20 minutes, respectively. Before the actual experiment, the participants were given 15 minutes to read through the background material and they received a 20 minute introduction to the learning environment. The collaborative argumentation phase lasted 80 minutes.

Results Concerning research question 1, the student group which was using scripts produced arguments of significant higher formal quality. With respect to research question 2, that is the effect on knowledge acquisition, students in the scripting conditions outperformed students from the non-scripting condition. Here, significant positive effects were found for both scripts with respect to their goal (guiding the construction of single arguments and guiding the construction of argument sequences, respectively). Concerning domain-specific knowledge, all participants improved significantly from pre-test to post-test. However, students in the scripted conditions did not improve significantly more than students in the non-scripted conditions. There were no interaction effects between scripts found on any of the results.

This study provides evidence that the use of structural scripts are indeed helpful to improve knowledge on argumentation structure. However, in order to strengthen these indications, long-term results should be considered. While it is likely that students remember things they just performed in a post-test, one cannot say that these effects will last and improve future argumentation. Since micro-scripts are designed to be internalized by the students, they are likely to be experienced as disturbing factor at some point. Then, a out-fading of these scripts may be beneficial (Bouyias et al., 2008).

In conclusion, system designers should be aware of the fact that they can direct student activities by means of specific interface design decisions such as ontologies or micro-scripts. These influences should be clear when designing an argumentation system in order to avoid potentially detrimental influences.

3.9.6 The Effect of Macro-Scripts

As mentioned previously, micro-scripts, in contrast to macro-scripts, usually do not make use of all the concepts that have been described as components of scripts in section 3.5. In order to evaluate the impact of the other components, this section deals with the concrete implementation of macro-scripts.

Scripting by Assigning Roles

Directly connected to the sequencing of activities is another concept that can be found in macro script definitions: the assignment of roles. These roles are intended to promote skills that are typical for certain activities, e.g., summarizing of important facts or moderating discussions about controversial topics (Webb & Palincsar, 1996). Here, the following study conducted by Schellens et al. (2007) evaluated the effect of role distribution in asynchronous collaboration settings.

Domain Learning and Instruction.

Hypotheses / Research Questions (1) What is the impact on student levels of knowledge construction and final exam scores of scripting the interaction in asynchronous discussion groups by assigning roles? (2) What is the differential impact on student levels of knowledge construction and final exam scores of being assigned a specific role?

Method Two successive cohorts of freshman students (n=223 and n=286) enrolled in an Instructional Sciences course at Ghent University participated in a field study. The course was taught by the same instructor and used the same course material. Participation was required in order to earn credit points. The students were divided into asynchronous discussion groups of approximately 10 students each. There were two conditions, a scripted one involving assignment of rotating roles (moderator, theoretician, summarizer and source searcher) during discussions and a non-scripted one without role assignments.

Results Students in the script condition showed a better performance than students in the no-script condition. The performance increase affects both the ongoing discussion processes as well as acquisition of domain-specific knowledge. However, the direct effect of single roles was controversial. Here, only the summarizer role resulted in significant improvements compared to the non-scripting condition concerning the level of knowledge construction during role assignment. The other roles did not improve. In fact, even a significantly negative effect was found for the source searcher role.

In conclusion, there seems to be a positive effect of assigning roles. However, the direct consequences which are implied by certain roles in different contexts (e.g., phase of the discussion or activity performed) are not clear, yet.

Supporting Different Argument Phases with Diagrams

The general purpose of scripts is to specify learning units which are supported by a combination of scripting concepts such as certain representations to scaffold specific learning activities. A study that evaluates the supporting role of graph-based argument representations for different activities is the one of Lund et al. (2007).

Domain Education.

Hypotheses / Research Questions To which extend does the use of argumentation graphs influence learning when used either (a) as medium of debate or (b) way of representing a chat debate?

Method 36 secondary school students participated in a 4 day quasi-experimental study. Before the actual experiment, the students were taught elementary notions of argumentations (Day 1). The experiment consisted of four phases including training (Day 2, 60 minutes), preparation (Day 3, 120 minutes), debate (Day 4, 70 minutes) and consolidation (Day 4, 30 minutes). The debate phase was the experimental condition. In condition one, the graph (in combination with a chat) was the medium of debate whereas in condition two, the graph was used to represent the debate carried out previously in the chat. The software tool used was DREW (Corbel et al., 2002).

Results The transformation of chat debates into a graph representation (condition two) led to a deeper conceptual understanding of the topic. Students used the graph as *"a unique voice"* (Lund et al., 2007, p. 292), that is there were less conflicting opinions present. However, in individual diagrams created after the intervention, they added significantly more non-argumentative relations, that is they add explanations and elaborations rather than supporting or opposing arguments. In contrast, the graph as medium of debate included more opinions instead of arguments.

These results show that the activity which is intended to be supported by a graph-based argument representation directly influences the outcomes that can be expected. By requiring additional transformation steps, as done here by transferring the content from the chat discussion to a graph-based representation, more elaborated arguments can be expected. An explanation is the additionally required reflection on the discussion. However, depending on the step in the argumentation process, it may be beneficial to cause opinion expression since a discussion of opposing views, for instance as part of a later step, is highly motivating. Thus, the context, which was mentioned frequently to be an important factor, can be designed to some degree by means of scripts.

Overall, one can conclude that the use of collaboration and learning scripts is beneficial to scaffold important parts of the argumentation process. Nevertheless, the extend of the scripting approach must be carefully evaluated. As reported by Dillenbourg (2002), there is a risk of *overscripting* present in all scripting approaches, that is micro- as well as macro-scripts, which could result in detrimental effects. However, recent research (Stegmann et al., 2011) argues that the fear of overscripting should not be pushed too far. Future research will have to evaluate the usefulness of various scripts to scaffold argumentation activities in order to provide guidelines how to improve learning further.

3.9.7 The Effect of Adaptive Support

An important part of learning is the recognition of mistakes. While teacher's time and availability is restricted, artificial intelligence techniques have been developed to assist student's learning.

LARGO

One of the argumentation systems involving analysis and feedback methods is LARGO. In order to evaluate the usefulness of these methods, there were basically two studies (Pinkwart et al., 2007, 2008) conducted. In this section, the results of the first one will be presented in detail. Concerning the second study, only additional insights and diverging results will be summarized.

Domain Legal argumentation.

Hypotheses / Research Questions Graphical format and advice would help students better identify and understand the argument components.

Analyses & Feedback Methods Structural analyses by means of a graph-grammar including prioritization of feedbacks as well as self-explanation prompts to encourage critical thinking. Feedback was given on-demand, that is students had to request feedback explicitly.

Method As part of a first year Legal Process course at the University of Pittsburgh, students were invited to participate in a lab study. Participation was voluntarily and paid with $80. The task was to work with extracts of the oral arguments from two personal jurisdiction cases. The lab study followed a between subject design with two conditions. In the control condition the students were required to highlight relevant passages and take notes using a notepad, whereas students in the experimental conditions represented the oral arguments graphically using the LARGO system and its feedback mechanisms. Overall, the experiment consisted of four sessions of two hours each distributed among a four week period following a pre-/posttest design using questions from the standardized Law School Admission Test (LSAT). 28 students finished the experiments.

Results Neither in the pre-test nor in the post-test scores were a significant differences found. A differentiation of participants by means of the aptitude levels represented by the LSAT scores into low, medium and high indicated that students with a low aptitude benefited more from the LARGO system than students with a higher aptitude when considering the near-transfer questions and questions regarding the evaluation of hypotheticals.

The results of the study are not clear. First of all, there was no significant effect found when considering all participants. Yet, a categorization into aptitude levels of students revealed further insights. However, the reason for the improvement is not completely clear. Still open is the question whether the argument visualization, the feedback or a combination of both was the reason behind it. Since there was no condition with LARGO but without the feedback mechanisms, no clear answer can be provided. A follow up post-hoc analysis, however, showed that the participants requested feedback at fair frequency. Moreover, this frequency increased over time which indicates that it was experienced helpful for the participants.

In order to replicate and enrich the results of the first study, Pinkwart et al. (2008) conducted a second study with LARGO. Instead of a lab study, a controlled field study was used. The main difference was that the use of LARGO was mandatory for first-year law students. Thus, the students were not paid for their participation. Further, the students' performances did not count for their course grade. In this case, the results could not be reproduced. Instead, there were significant negative effects found considering the post-test of the experimental condition using LARGO. An in-depth analysis of the results highlighted that the students refute the use of the feedback mechanisms. In fact, the amount of feedback requests dropped by 80% compared to the first study and even decreased over time. This difference could potentially have caused the decrease, which would imply that the graph grammar based feedback is indeed helpful for improvements. However, the researchers hypothesized that the motivation of the students could have been a key factor in this setting. Compared to the first study, no concrete benefits neither in form of financial reward nor in form of grades were offered in the second study. This matches with other researchers results which indicate that engagement and, hence, motivation is directly related to overall success (Schellens et al., 2005; Munneke et al., 2007).

In addition to the feedback mechanisms evaluated in these studies, LARGO comprises a collaborative filtering approach. Based on internal references to the transcript of the oral arguments used in LARGO, the assessment of users who have been working on the same part of the argument will be used to assess other users solutions. However, this peer review approach has not been evaluated in LARGO, but in other (non-argumentation) learning contexts (Loll & Pinkwart, 2009a,b,c). In (Loll & Pinkwart, 2009c) a lab study confirmed that the LARGO algorithm is able to provide acceptable quality heuristics once 3-5 user assessments have been collected. Further, the algorithm has been applied successfully as means for university courses at Clausthal University of Technology as part of the eLearning system CITUC (Loll & Pinkwart, 2009a,b).

Convince Me

Another argumentation system with analyses and feedback mechanisms is Convince Me (Schank, 1995; Schank & Ranney, 1995). Instead of directly assessing the argument structure, Convince Me aims to promote coherent argumentation by checking for internal consistency of user believability/acceptability assignments.

Domain Scientific argumentation.

Hypotheses / Questions Does Convince Me make its users better reasoners?

Analyses & Feedback Methods Structural analysis by means of the ECHO algorithm. The ECHO algorithm is a concrete connectionist implementation of the Theory of Explanatory Coherence (TEC) (Schank, 1995). It checks for internal consistency of the argument by means of correlation between user-produced argument structures and their self-assignment belief ratings of the argument structures.

Method Twenty undergraduate students with various backgrounds went through a pre-test, three curriculum units on scientific reasoning, integrative exercises from multiple domains (biology, medicine, ethics, and physics) with two competing theories each, a post-test and a final questionnaire. They were split up into two groups with ten students each. One group, that is the experimental one, was using Convince Me for the three curriculum units, while the other group, the control one, used paper and pencil for them. The pre- and post-test were done on paper, that is without Convince Me for both groups.

Results The results revealed that the experimental group performed superior compared to the control group: The users' belief-activation correlation was significantly higher than the correlations of their own pre-test and higher than those of the control group. Furthermore, it did not drop significantly during the post-test in which no tool was available, which indicates transfer effects. The control group, however, did not show any significant improvements between pre- and post-test.

The results of the Convince Me study showed a significant gain between pre- and post-test. However, similar to the LARGO study, there was no control condition which used the tool without feedback. Thus, both studies compared the use of a tool including argument visualizations, ontologies and feedback to a control condition in which none of these factors were present. In essence, one can conclude that the context, once again, is an important factor that needs to be considered. Even though there are promising approaches present to provide feedback in the complex, ill-defined field of argumentation (see chapter 2.3), the future will show whether they are accepted by the learner or whether classic teacher oriented feedbacks are still favored.

3.10 Conclusion & Outlook

Assisting students in their acquisition of argumentation skills is an important educational goal. This has been widely recognized by researchers and educational communities from various disciplines including science, the law and ethics. The presented review of argumentation systems and methods developed in the last (roughly) 15 to 20 years revealed a huge variety of ways to assist argumentation and empirical evaluations of the approaches highlighted potential benefits and lacks.

Even though achievements have been clearly made in the area of computer-based argumentation, it might be harder than ever to design a reasonable argumentative learning activity. The reasons for that are manifold. On the one hand, one is spoilt for choice. This includes the choice of an adequate argument representation and ontology, an appropriate interaction and process design as well as an effective analysis and feedback method. On the other hand, the curse of science does not draw back from the area of computer-based argumentation: Each achievement exposes new questions. Thus, the question whether the area of computer-based argumentation has been fully explored can be clearly answered: *No.* Instead, there is quite a large set of open challenges that need to be solved in the future.

The first challenge is the absence of an overarching theory how exactly computers can support the education of argumentation skills. Even though, there are some empirical evaluations present, the review revealed partly conflicting or even contradictory results. Examples include the role of argument mapping and the degree of granularity of ontologies. The main problem here is that the results are hardly comparable because single factors are rarely evaluated. Instead, the use of multiple tools in different contexts manipulates an unmanageable amount of influencing factors that are likely to have major impacts on the results. Thus, more systematic empirical research is required in order to create a theoretical model of the computer's role in educating argumentation skills.

The second challenge is somewhat connected to the first one. As mentioned before, the variability of approaches is imposing. However, only few of them have been subject of empirical research. Thus, the direct consequences with respect to educational benefit of certain design decisions are largely unclear. Just imagine the design of an argumentation ontology. As shown in section 3.6, the amount of different argumentation ontologies is nearly as large as the number of available argumentation systems. While the differences are especially obvious when considering general-purpose systems, there are even differences within the same argumentation domains. In addition, the literature reports controversial findings. Whereas Suthers (2003) reported that a complex ontology confuses students, other researchers (Jeong, 2003; Soller, 2001, 2004) reported that their students did not have any problems dealing with it. The simple reason for this kind of absence of research is that these experiments are hardly doable. In fact, existing argumentation systems are not flexible enough to be used as tool for research.

The third challenge is the premature commitment to design decisions and the ignorance of important concepts. An example for the first part is the favored use of graph-based visualizations. The empirical findings so far did not provide forcing evidence that graph-based argument representations are best for educational purposes. Instead, other approaches have been shown to be effective within scripted scenarios. In combination with the previously stated challenges, the role of design decisions should be carefully evaluated and requires further studies. Considering the second part of the challenge, important aspects of everyday argumentation have been embezzled. These include, for instance, facial expressions, gestures and tone of speech which are highly important for face-to-face argumentation (Roth, 2000; Allwood, 2002; Lund, 2007).

The fourth challenge is the role of collaboration. In fact, a considerable number of argumentation tools in the review did not involve any support for collaboration. The reason though is clearly the missing technological capabilities available when the research field evolved. However, first empirical results indicate that the involvement of collaboration, while beneficial on the one hand, is causing additional problems that need to be solved on the other hand. Thus, additional research in this direction is clearly required.

The fifth challenge is connected to the technology side of argumentation systems. Here, the review revealed an almost total lack of system documentation and research publications about generic, flexible, and reusable system design patterns for building these systems. Not only do, apparently, few people conduct research on educational argumentation systems from a computer science perspective, but also the existing tools are not well described from a technical viewpoint. This unfortunate situation imposes severe implications on the research community: Whenever a researcher wants to design an argumentation system in order to evaluate a hypothesis, the wheel has to be reinvented again and again. This results in considerable effort that could be avoided.

The sixth and final challenge is the question how to motivate the use of involve argumentation systems in actual classrooms in order to allow students to benefit from it. Here, basically two barriers must be taken: The first one is the technological barrier. Whereas existing argumentation systems are usually desktop applications that require a manual installation on each machine, recent and future argumentation systems should eliminate this annoying step, for instance, by using a web-based approach. The second one is the social barrier. Here, the needs of people involved with the tools' use, for instance, teachers and students, must be considered. Therefore, they should be involved in the design process of the system in order to convince them of the benefits provided by the tools.

In this thesis, a first step to face these open challenges will be taken. In order to get an overarching theory of how to support the education of argumentation skills via computers (challenge 1), this thesis will focus on the prerequisites to obtain such a theory. These prerequisites can be found in the other challenges. Here, three directions have to be considered: (1) The *lack of clarity* concerning the effects of various approaches (challenges 2 & 4) in combination with premature commitment to design decisions (challenges 3), (2) the *lacks on the technology side* (challenges 4 & 5) as well as (3) the *motivation* to actually use argumentation systems in education (challenge 6). In this context, the technology side is a key point. On the one hand, it is essential to conduct valid and comparable research to achieve more in-depth insights into the impacts of certain design decisions. On the other hand, the technology must be experienced as support instead of an obstacle, that is, it must provide a good user experience to motivate people to use it. Thus, this thesis will concentrate on the fifth challenge.

The central question of this thesis *(How can computers provide domain-independent support for the education of argumentation skills?)* is directly related to the fifth challenge and was split into five sub-questions. The review provided first answers to sub-questions 1 *(What are the differences and similarities of argumentation in different domains and how may these influence the design of tools that support argumentation?)*

and 2 *(Which practices have shown to be effective in existing approaches to scaffold argumentation, which failed and what are the reasons for it?)*. These answers, even though not exhaustive, can be seen as starting points to face challenges one to four. More important is, however, that these answers will be used in the next chapters to answer sub-questions three to five and, hence, will provide a significant contribution to the fifth challenge.

Before proposing an actual solution for the technological issues, the answers provided in this chapter will be extended. Thus, the next chapter will extend the review by means of a survey among argumentation experts.

4 Filling the Gap to Current Developments: A Survey among Experts

The review of existing argumentation systems and methods presented in the previous chapter provided a detailed overview of prior work. On the one hand, it revealed an impressive variety of approaches and highlighted promising ways to educate argumentation skills. On the other hand, it showed that concrete answers to important questions are still missing.

A reason for the latter may be found in the way a literature review works. Based on published work, including descriptions and evaluations of existing approaches, an overview of the field is created. While this is an adequate way to get insights into research fields in which exhaustive work has been done and concrete answers what is good and what is not are present, this is only partly true for younger fields such as computer-supported argumentation in which these answers are not present yet. As a matter of fact, reviews in these fields are limited in several ways. First of all, a review only captures the past. In an immature field such as computer-supported argumentation the direct consequence is that new developments that are currently on their way will not be covered. Second, the motivation that led to certain design decisions will not be revealed. Thus, the context is not caught adequately. An example is the role of collaboration in existing argumentation systems — is it really technological limitations that led to absence of collaboration or was it on purpose? Third, failures are rarely reported. This implicates other researchers will make the same mistakes over and over again. In addition, promising ideas that did not result in direct success may perish. Finally, publications are usually pressed for space and, hence, biased towards results instead of documentation.

Together, these limitations, which raises additional questions, and the still open questions motivated an extension of the review. By means of a survey among argumentation experts from different argumentation domains, this gap between published work and actual domain knowledge present in the integrated experiences of experts will be closed. Therefore, this chapter will deepen the preliminary results published in (Loll et al., 2010a,b, 2011a) by means of additional results in order to extend the answers to research questions one to three (see chapter 1.2).

4.1 Method

4.1.1 Research Questions

To extend the review (see previous chapter; Scheuer et al. (2010)) appropriately, the underlying research questions of the survey were motivated as following: On the one hand, the results of the review should be consolidated and the motivation behind specific design decisions such as visualizations or feedback strategies and timings should be uncovered. On the other hand, the results of the review should be deepened and extended, that is additional insights into computer-based argumentation should be revealed such as important lessons learned from past developments and evaluations.

Therefore, the research questions are divided into multiple categories. The first category deals with the role of computer-based tools to support the general argumentation process as well as the impact of computer-based tools on argumentation learning. In particular, this question will be used to get a more in-depth understanding of which processes are present in argumentation that is applied in practical learning and decision finding processes. In the second category, the role of collaboration is evaluated. Here, it should be clarified when different collaboration settings are appropriate. In addition, typical errors that occur in individual and collaborative argumentation will be collected. The third category deals with questions concerning the (automatic) evaluation of arguments and the provision of adequate feedback. The final category will collect possible guidelines that should be incorporated into future developments to benefit from empirical findings as well as from experts' experiences. This way, potential usability lacks that may hinder the use of argumentation tools in practice can be revealed and avoided in future.

For each research question, there are primary and secondary survey questions reported. The primary survey questions indicate which survey questions were designed to answer this research question whereas the secondary survey question were considered to be useful to support the general findings for instance via correlation analyses. All questions can be found in Appendix B. The answers to the open-ended survey questions will be backed up with direct quotes from the experts. However, the experts were ensured to stay anonymous. Therefore, there will be no author added to the quotes.

Computer-support for the Argumentation Processes

(RQ1) **How does a typical argumentation process look like?**

In section 3.5 of the review the use of collaboration scripts was described. These scripts were designed to scaffold specific parts of the argumentation process. However, to identify potential alternative application areas it is important to get an overview of a typical argumentation process that is applied in current approaches to teach argumentation. This way, a set of activities that can be supported within scripted approaches will be collected to identify potential starting points for future research.

Primary survey questions: Open-ended question (Appendix B.3) 1c

(RQ2) **Can computer-supported / computer-mediated argumentation replace face-to-face argumentation?**

Even though there is a vast set of computer-based argumentation tools available (see Appendix A), it was not clear in the review what their actual purpose is. Imaginable scenarios here include (a) the computer as addition or support to classic face-to-face argumentation methods and (b) the computer as replacement of classic face-to-face methods. However, these scenarios come along with unequal requirements and demands toward the computer system and its users: Scenario (a) requires the computer system to be able to actually assist with classic teaching methods whereas the demands towards the capabilities of the computer system would grow rapidly in scenario (b). Thus, it is important to clarify what the users expect from an argumentation system before actually designing or implementing one in order to make a step forward toward an overarching theory how computers can support argumentation.

Primary survey questions: Multiple choice questions (Appendix B.2) 4 & 12

Secondary survey questions: Personal background question (Appendix B.1) 4

(RQ3) **Are visual representations helpful for learning and / or understanding argumentation?**

The approaches to visualize abstract arguments by means of graphical representations have been described in section 3.2. In addition to the evaluations presented in sections 3.9.1 & 3.9.2, this question aims to catch the experts' opinions regarding these visualizations. This way, further insights into specific strengths and weaknesses might be revealed and the motivation behind the decisions that led to the use of argument visualizations will get clear.

Primary survey questions: Multiple choice questions (Appendix B.2) 3 & 11, open-ended question (Appendix B.3) 1b

Secondary survey questions: Multiple choice question (Appendix B.2) 12

Individual & Collaborative Argumentation

(RQ4) **What type of collaboration is appropriate for learning to argue?**

In practice, argumentation is used to solve complex problems (see section 2.3). Usually, multiple points of views are involved and the decision making process consists of discussions among groups. However, the review revealed that (especially early) argumentation systems are designed to support only individual argumentation. Therefore, this question should clarify the open issue if there are application areas or argumentation domains in which individual argumentation is more suitable than collaborative argumentation or if there are scenarios in which variations or combinations of these settings are more appropriate.

Primary survey questions: Multiple choice questions (Appendix B.2) 2 & 6, open-ended question (Appendix B.3) 1c

(RQ5) **Does the formality of a domain influence the type of collaboration that is appropriate?**

As mentioned in section 2.3, argumentation is a highly complex endeavor. In addition, collaboration and structured processes as described in section 3.5 may put additional efforts on the users who might get overwhelmed by cognitive demands. Thus, this question is designed to check for a possible interaction effect between the formality of a domain and the type of collaboration. Imaginable here is for instance that an increased degree of formality might be supported by means of formal restrictions which allow establishing common rules. On the other hand, formal restrictions may even hinder a fluent collaboration process.

Primary survey questions: Multiple choice questions (Appendix B.2) 1, 2, 5 & 6

Secondary survey questions: Multiple choice question (Appendix B.2) 12

(RQ6) **What are typical mistakes that occur in argumentation and do they differ in individual and collaborative argumentation settings?**

One of the most important steps in a learning process is the recognition of mistakes. Once you understand what you have done wrong, you should be able to correct your mistakes and avoid them next time. Therefore, it is — especially from a learning to argue perspective (see chapter 2.2.1) — important to identify potential error sources. This research question aims, in particular, at revealing errors that are typical for certain collaboration settings. This way it may get clear what kind of support is beneficial for individual and group learning.

Primary survey questions: Open-ended question (Appendix B.3) 2a & 2b

Analysis & Feedback

(RQ7) **What criteria are applied in practice to determine whether an argument is good?**

Whereas giving feedback is an important step in the learning process, it is especially hard in ill-defined domains such as argumentation (see chapter 2.3) in which multiple solutions are imaginable and one usually cannot judge between right and wrong, but good or bad. Thus, this research question aims to collect criteria that are applied in practice to judge the overall quality of arguments.

Primary survey questions: Open-ended question (Appendix B.3) 1a

(RQ8) **Is it possible to develop automated analysis features that can effectively analyze arguments and are there domain-specific differences?**

In the introduction of this thesis, the workload of teachers as well as the time consuming process of teaching argumentation has been highlighted. Therefore, alternative concepts to provide help to learners are of major interest. This research question will evaluate if the experts believe that computers and their limited analytical capabilities (see chapter 3.7) are able to analyze arguments as first step of the feedback process.

Further, it is also important to check if there are domain-specific differences that may affect the suitability of automated analysis. In highly structured domains, for instance, it might be easier to actually provide analysis tools as the review of existing analysis and feedback methods indicated (see sections 3.7 and 3.9.7).

Primary survey questions: Multiple choice questions (Appendix B.2) 7, 8 & 9

Secondary survey questions: Personal background question (Appendix B.2) 2

(RQ9) **Who is going to provide tutorial feedback in practical argumentation learning settings and when is it most effectively provided?**

Once adequate analyses, either from a teacher or an artificial intelligence agent, has been conducted and possible feedbacks are generated, the next question is how and when the feedback should be provided to the learner. In chapter 3.7 multiple strategies and timings have been discussed. However, the evaluations of these techniques (see chapter 3.9.7) were rare and provided ambivalent results. Thus, the experts' opinions are of interest to clarify which of these approaches are favored in practice and why.

Primary survey questions: Multiple choice question (Appendix B.2) 10, open-ended question (Appendix B.3) 2c

Secondary survey questions: Multiple choice questions (Appendix B.2) 1, 6, 7, 8, 9, 11 & 12, personal background question (Appendix B.1) 5

Guidelines for Future Developments

(RQ10) **What kind of flexibility is present in existing argumentation technology to face the challenges of various argumentation settings and domain-specific differences?**

The previous chapters described the aspects in which argumentation in various domains differs. The review of existing argumentation technologies showed that there are domain-specific systems as well as systems that aim to provide rather general support for multiple applications. In addition, the review revealed the need for a flexible argumentation system to enable more systematic research (see chapter 3.10). Fortunately, the idea to generalize support grew in the last years. Thus, this question is designed to actually collect current development approaches to provide flexibility in computer-based argumentation support.

Primary survey questions: Open-ended question (Appendix B.3) 3b

(RQ11) **Are there any guidelines that should be complied for future developments based on the experts' experiences?**

Literature is usually biased to stories of success. Thus, a literature review is not capable of collecting failures. Nevertheless, these *lessons learned* are invaluable for future developments and research. Based on these lessons learned, this research question's goal is to identify potential guidelines for the future in order to motivate the use of argumentation systems in research as well as in teaching.

Primary survey questions: Open-ended questions (Appendix B.3) 3a & 3c

4.1.2 Selection of Participants

To select an adequate audience of the survey, a list of approximately 40 people was created. This list contained experts that were personally known to the LASAD project group as well as developers of argumentation systems investigated in the review (see section 3). Then, the list was extended by means of a systematical search through the author lists of relevant conferences (ITS[1], AIED[2], CSCL[3], COMMA[4], ISSA[5]) and journals (ijCSCL[6], IJAIED[7]) as well as multiple Google (Scholar) searches. For individuals that were unknown to the project team, their home pages and publication lists were checked to ensure expertise in (computer-supported) argumentation. In total, a compiled list of 153 experts was created. All experts were personally invited via e-mail by the project team leaders to take part in the web-based survey.

Participation was voluntary. As motivation an Apple iPod was raffled among all participants and the anonymized results of the questionnaire were offered.

4.1.3 Design of the Questionnaire

The survey consisted of four steps. In the first step, participants were informed about the purpose of the survey, the use of the data and were asked whether they wanted to receive the anonymized results of the survey after analysis. Second, they were asked

[1] International Conference on Intelligent Tutoring Systems
[2] International Conference on Artificial Intelligence in Education, http://iaied.org/
[3] International Conference on Computer-Supported Collaborative Learning, http://www.isls.org/conferences.html
[4] International Conference on Computational Models of Argument, http://www.comma-conf.org/
[5] International Conference on Argumentation of the International Society for the Study of Argumentation, http://cf.hum.uva.nl/issa/
[6] International Journal of Computer-Supported Collaborative Learning, http://ijcscl.org/
[7] International Journal of Artificial Intelligence in Education, http://ijaied.org/

for their professional background (see Appendix B.1). All participants with a self-reported experience of greater or equal to 3 (*somewhat*) on a 5pt Likert scale in at least one of the categories (*research*, *teaching* and *designing/developing of argumentation systems*) were then, in the third step, asked the multiple-choice questions listed in Appendix B.2. Here, there were usually two questions designed to measure the same thing, e.g. question 3 (*"I think that an argument represented in visual fashion — for instance, as a graph of participants' points and their relationships — can be very helpful to people in understanding and reflecting upon the argument"*) and question 11 (*"Arguments shown in graphical fashion are likely to be helpful in learning or understanding argumentation"*). This way, it was possible to test for internal consistency of the responses and to increase the robustness of the questionnaire with respect to slight formulation weaknesses that may occur in non-standardized surveys. Finally, there were the open-ended questions listed in Appendix B.3. Here, the experts were assigned to up to two categories (*research*, *teaching*, *development*) in which they had a self-reported expertise of greater or equal to 3 (*somewhat*). If the self-reported expertise in two or more categories was equal and above the threshold, they got distributed randomly to the respective conditions.

4.2 Results

4.2.1 Coding of Open-ended Questions

To evaluate the open-ended questions, the LASAD project team created taxonomies (hierarchical coding trees; see Appendix C) to represent the answers given to each of the questions. The taxonomies allowed the grouping of similar answers while, at the same time, highlighting the most frequently occurring categories and answers. The taxonomies were created in multiple steps as follows:

1. Initial taxonomies were created by having two raters independently code a randomly selected set of 20% of the responses. In this first round, the first rater coded 5 of the 9 open-ended questions; the other rater took the remaining 4 questions.
2. Next, another 20% of each question's responses were randomly selected, with the raters switching the questions to code (that is, rater A evaluated rater B's questions from round 1, with rater A developing new taxonomies for those questions, and vice versa). The resulting taxonomies from rounds 1 and 2 were merged, with conflicts solved by discussion.
3. Finally, another 10% of the responses, again randomly selected, were taken and coded by the two raters independently using the existing taxonomies.

   The inter-rater reliabilities of the last 10% were calculated using percentaged agreement, which resulted in the high values presented in Table 4.1 (Note: Measures such as Cohen's κ are not applicable due to the hierarchical nature of the coding scheme). Here, the agreement on category level means that both raters assigned at least the same top-level category, whereas agreement on details level means that both raters assigned the same sub-level category. A single response was often assigned to more than one category, since the lengths of the answers were not restricted and answers usually contained multiple parts. Again, conflicts were solved by discussion.

Question	Number of Answers	Agreement (Category)	Agreement (Detail)
RQ1	$n = 59$	90.5%	76.2%
RQ2	$n = 58$	71.4%	64.2%
RQ3	$n = 42$	80.0%	75.0%
TQ1	$n = 40$	100.0%	90.9%
TQ2	$n = 29$	77.8%	77.8%
TQ3	$n = 37$	80.0%	80.0%
DQ1	$n = 59$	65.2%	60.9%
DQ2	$n = 48$	84.2%	84.2%
DQ3	$n = 58$	89.5%	63.1%

Table 4.1: Inter-rater reliability using percentaged agreement for coding of responses to open-ended questions

Overall, the agreement was assumed to be high enough on both the category as well as the details level.

4. Having established an agreed-upon coding scheme, the remaining 50% of the survey answers were distributed among the raters and categorized individually. The resulting taxonomies can be found in Appendix C.

4.2.2 Participants

In the two month the questionnaire was online, 97 of 153 experts provided usable responses. Usable answers excluded all responses with a self-reported expertise of less than 3 (*somewhat*) in all areas (*research, teaching, development*). The answers of participants who reported multiple domains as their primary domain of expertise were counted in all of them. The final list of filtered responses contained 34 from science, 8 from law, 16 from education, 3 from ethics, 3 from business, and 23 from another argumentation domain including medicine, philosophy, business, computer science, mathematics and politics.

Most of the experts have been using computer-based argumentation tools before as indicated by the medians[8] presented in Figure 4.1. A non-parametric Kruskal-Wallis

[8] Medians are reported, because the 5pt Likert scale is ordinal, that is one cannot assume the differences between all possibilities are equals. Thus, means and standard deviations would not have been appropriate. In the rest of this chapter the abbreviation m stands for median. In addition, the error indicators show the interquartile range, that is the bottom limit indicates the first

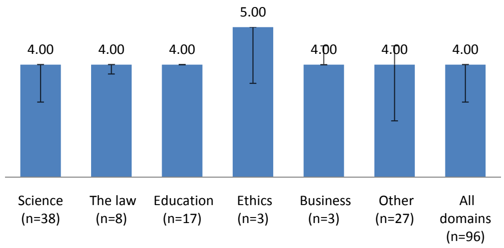


Figure 4.1: Self-reported familiarity with computer-based argumentation tools among argumentation experts. [Personal background question 4: I have used computer tools that support argumentation (Scale: 1=Very little, 2=Little, 3=Somewhat, 4=A lot, 5=I am an expert)]

test[9] did not show any significant differences between argumentation domains ($H(3) = 0.316, p = 0.957$).

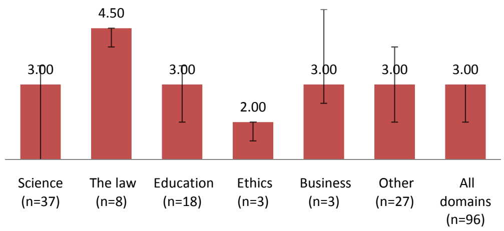


Figure 4.2: Self-reported knowledge of artificial intelligence technology among argumentation experts. [Personal background question 5: I have knowledge of artificial intelligence technology (Scale: 1=Very little, 2=Little, 3=Somewhat, 4=A lot, 5=I am an expert)]

Concerning the experts' knowledge of artificial intelligence (AI) technology, a non-parametric Kruskal-Wallis test pointed out that experts in the area of legal argumentation are significantly more familiar with AI technology ($m = 4.50$) than experts from other argumentation domains ($H(3) = 13.755, p = 0.003$ with a mean rank of 44.64 for "*Science*", 80.25 for "*The law*", 39.33 for "*Education*" and 50.14 for "*Other*"). The resulting medians are shown in Figure 4.2.

quartile whereas the top limit indicates the third quartile.

[9] A non-parametric Kruskal-Wallis test was used instead of an ANOVA because the 5pt Likert scale is ordinal, that is one cannot assume the differences between all possibilities are equals. Whereas this point is still discussed in the literature (Jamieson, 2004), I will follow the strict recommendations for ordinal scales. In this concrete case and all following questions in which a Kruskal-Wallis test is performed, the categories with only few answers — that is "*Ethics*" and "*Business*" — were merged into the "*Other*" category to get more reliable results.

4.2.3 Computer-support for the Argumentation Processes

Research question 1: How does a typical argumentation process look like?

Argumentation involves a complex process. This is reflected in the experts' answers to the open-ended question *"Can you describe the typical processes and roles that are used in argumentation (or debates between parties) in your primary domain of interest?"* shown in the taxonomy presented in Figure C.3. Here, the experts broke argumentation down into basically four phases: (1) the preparation phase, (2) the discussion phase, (3) the decision phase and (4) the result phase. Each of these phases consists of a set of activities which are collected in Figure 4.3. However, the resulting collection of activities in the process, which is shown in Figure 4.3, was not mentioned by a single expert. Instead, most experts mentioned parts of it.

As described in the review (see chapters 3.5, 3.9.5, and 3.9.6) the argumentation process is likely to involve roles to scaffold certain argumentation activities. In the survey, the experts prefer to describe tasks which are connected to roles. Typical tasks mentioned by the experts include: (1) a specific position is assigned to one party, (2) each party takes its own position, (3) one participant or a group of participants states a position and others criticize and modify this position. These tasks are sometimes rotated between participants. Obviously, most experts assumed arguing in groups. Individual argumentation processes was only mentioned by a minority of experts.

Surprisingly, a considerable amount of experts (25%) did either not report any processes and roles or did not understand what was meant by the question which indicates that concrete process specifications are not an integral part of argumentation in all domains and application areas.

Research question 2: Can computer-supported / computer-mediated argumentation replace face-to-face argumentation?

As Figure 4.4 indicates most experts agree that computer systems have the potential to support and even improve argumentation among spatially distributed people ($m_{Q12} = 4.00$). However, the slightly more strict formulation of survey question 4 which implied that computers can replace classic face-to-face methods is less clearly answered ($m_{Q4} = 3.00$). Whereas experts from an educational background agree that it is possible to relinquish face-to-face settings, this point is seen controversial in many other domains such as ethics, business, the law and science. The concrete role of the computer-based tool is, hence, dependent on the domain in which argumentation takes place. In addition, the experience with computer-based argumentation tools correlates[10] with the agreement that computers can either improve or even replace face-to-face argumentation ($\rho_{ToolsExp.,Q4} = 0.217, p = 0.037, n = 92;$

[10]In this complete chapter, Spearman's rank correlation was used when correlation is mentioned, because of the ordinal Likert scales used in the survey.

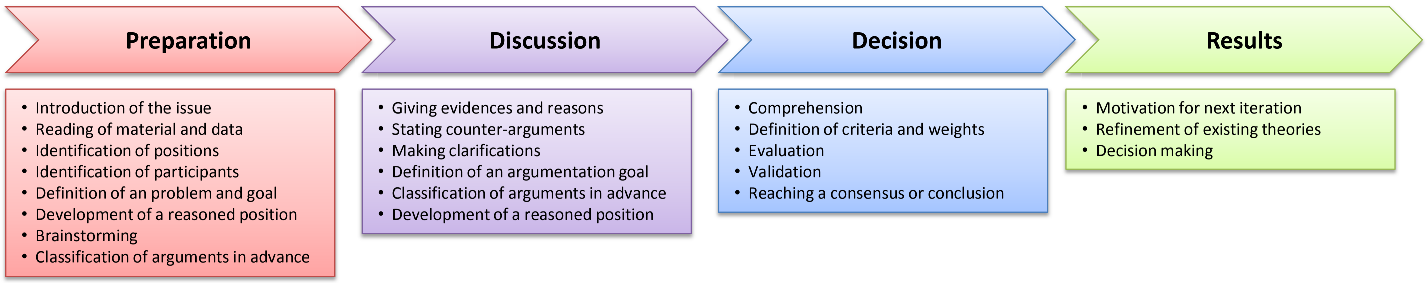
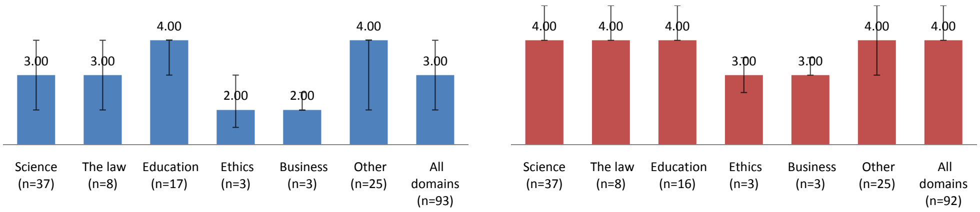


Figure 4.3: Integrated expert model of an argumentation process



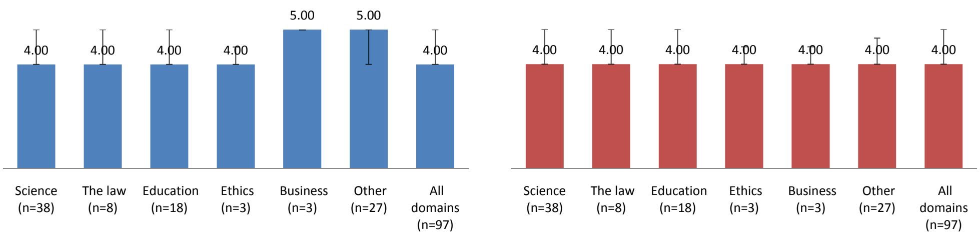
(a) Question 4: The kind of argumentation I am interested in can be facilitated or supported by a computer system well enough so that the participants do not need to speak and argue face-to-face.

(b) Question 12: Computer systems have the potential to support people in conducting useful, valid arguments over the Internet, perhaps even improving upon standard, face-to-face discussion.

Figure 4.4: Evaluation of the computer's role as mediator or replacement of face-to-face argumentation (Scale: 1=Strongly disagree, 2=Disagree, 3=Neither agree nor disagree, 4=Agree, 5=Strongly agree). Correlation between questions: $\rho = 0.435, p < 0.001, n = 90$

$\rho_{ToolsExp.,Q12} = 0.215, p = 0.039, n = 92$). This might indicate that the confidence in computer systems for argumentation is dependent on the experience with these tools. Thus, reservations might get reduced once the users get familiar with these tools and their possibilities.

Research question 3: Are visual representations helpful for learning and / or understanding argumentation?



(a) Question 3: I think that an argument represented in visual fashion — for instance, as a graph of participants' points and their relationships — can be very helpful to people in understanding and reflecting upon the argument.

(b) Question 11: Arguments shown in graphical fashion are likely to be helpful in learning or understanding argumentation.

Figure 4.5: Evaluation of the role of graphical argument visualizations (Scale: 1=Strongly disagree, 2=Disagree, 3=Neither agree nor disagree, 4=Agree, 5=Strongly agree). Correlation between questions: $\rho = 0.548, p < 0.001, n = 96$

The results of survey questions 3 and 11 shown in Figure 4.5 highlight that the vast majority of experts agree that graphical argument visualizations are helpful for understanding, reflection and learning argumentation. This is independent from the concrete domain in which the argumentation takes place ($H_{Q3}(3) = 3.689, p = 0.297$; $H_{Q11}(3) = 0.672, p = 0.880$).

Further, there is a moderate but significant correlation between questions 3 and 12 (*"Computer systems have the potential to support people in conducting useful, valid arguments over the Internet, perhaps even improving upon standard, face-to-face discussion."*) ($\rho = 0.291, p = 0.005, n = 92$) as well as between questions 11 and 12 ($\rho = 0.297, p = 0.004, n = 92$). Thus, the visualization that is provided by computer systems might be one of the reasons why experts think that computer support can even improve face-to-face argumentation.

However, even more interesting are the reasons why experts think that argument visualizations are helpful. Therefore, the answers to open-ended question *"Imagine that you have a software tool with graphical components representing different plausible argument moves that users can choose from. They might be able to choose from components such as claim, fact, or rebuttal and then fill in the selected shapes with text specific to their idea. Do you think such an approach would help or hinder users as they construct arguments and why?"*) are worth a closer look. The resulting taxonomy of answers shown in Figure C.2 highlights that visual argument representation are not helpful per se in any case. Instead, the majority of experts agree that they have advantages as well as disadvantages. On the advantages' side, the two most important things mentioned are: (1) visualizations are representational aids which especially help organizing and (2) they support learning. The following quotes back these points up:

> *"[Visual representations of arguments] should help users separate arguments into their essential component parts, and allow them to analyze and judge them accordingly."*

> *"These types of structures can help focus and refine student interaction. For example, the structures might help students better understand what appropriate types of arguments to make (e.g. "I should probably critique this now..."). There must be a balance between guiding the process in which a student engages, and allowing freedom and ease-of-use that will stimulate and promote learning."*

> *"[Visual representations of arguments] will help students to become aware of the functions of items like pro argument, con argument, support, fact and rebuttal in an argumentation [. . .] and it will help them to discover 'holes' in the argumentation like only specifying pro arguments and not refuting con arguments"*

On the cons' side the artificial approach of argument visualizations was mentioned as well as the limited expressiveness that is caused mainly by fixed pre-defined shapes. Typical quotes here included:

> *"[Visual representations of arguments] could also be regarded as highly artificial since considering these argument elements and explicating them is very atypical and unnecessary in everyday argumentation."*

> *"[Visual representations of arguments] might stop the proce[s]s of students' thinking during the discussion as they are busy in constructing of response to con[t]ributions"*

> *"There may be some trade-off in the degree of expressiveness afforded depending upon how strict the underlying argument structure is."*

Taking both sides into account, one can clearly say that the potential benefit of argument visualizations depend on the context in which they are applied. Depending factors that had been mentioned by the experts include the pedagogical goal, the amount of experience and training of users as well as their attitude and background. Further, the capabilities of the software directly influence its appropriateness to support argumentation, as revealed by the following quotes:

> *"It depends! Novice learners may benefit from the instructional support while experienced learners may be hindered to apply their advanced strategies"*

> *"There is no such general ontology for all domains and cultures. In addition, your categories "eg rebuttal" come from a specific theory. The best is to have a generic tool, work on the relevant structures and argumentative metalanguage in that domain."*

> *"It depends on the nature of the arguments and the training and attitude of the users."*

> *"The ontology of your labels greatly influences the categories in which people think. Hence, if you want to give people experience with using such categories it may help. In argumentation for learning this is only of limited value, much more matters the extent to which the argument advances the process "*

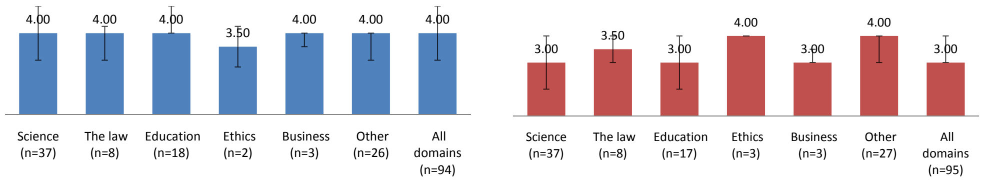
4.2.4 Individual & Collaborative Argumentation

Research question 4: What type of collaboration is appropriate for learning to argue?

Figure 4.6 shows that discussions are the preferred method to learn reasonable argumentation. Even though the experts do not exclude individual study and practice as way to learn argumentation (see question 6), most of them agree that collaboration is essential for learning purposes (see question 2). This is also supported by the taxonomy of the responses to the open-ended question *"Can you describe the typical processes and roles that are used in argumentation (or debates between parties) in your primary domain of interest?"* (shown in Figure C.3 in the Appendix) where nearly 100% of the experts agreed that the typical argumentation process in their domain include multiple parties and, hence, collaborative settings. This way it is expected to bring multiple perspectives together as the following quotes show:

> *"Argument[ation] is a group activity."*

> *"In situations involving multiple parties each will generally bring a different perspective."*

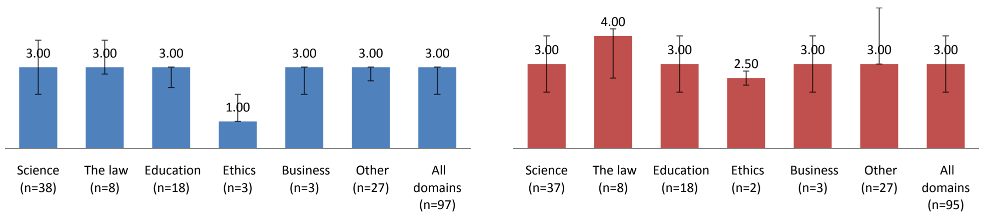


(a) Question 2: In my primary domain of interest, it is important that people learn argumentation through discussions, rather than on their own (e.g., from a book, by sketching arguments on paper, etc.).

(b) Question 6: The rules and valid forms of argumentation can be learned in my domain of interest through individual study and practice.

Figure 4.6: Evaluation of different types of collaboration for argumentation learning (Scale: 1=Strongly disagree, 2=Disagree, 3=Neither agree nor disagree, 4=Agree, 5=Strongly agree). Correlation between questions: $\rho = -0.182, p = 0.082, n = 92$

Research question 5: Does the formality of a domain influence the type of collaboration that is appropriate?



(a) Question 1: Which answer best describes the type of argumentation that is taught and/or used in your primary domain of interest? (Scale: 1=Informal, 2=Somewhat Informal, 3=Partly Informal / Partly Formal, 4=Somewhat Formal, 5=Formal)

(b) Question 5: To what extend do rules of inference and/or axioms provide structure to arguments in your primary domain of interest? (Scale: 1=Very little, 2=A little, 3=Somewhat, 4=A lot, 5=A very great deal)

Figure 4.7: Evaluation of formality of argumentation domains. Correlation between questions: $\rho = 0.356, p < 0.001, n = 95$

As shown in Figure 4.7, argumentation in general includes formal as well as informal aspects in the experts' minds. This seems to be also true for domains which are assumed to be rather formal such as mathematics (counted as "*Other*"). Rare exceptions are the law in which rules of inference and/or axioms provide a lot of structure (see question 5) and ethical argumentation which heavily depends on informal argumentation methods (see question 1). With respect to the research question, there was no significant correlation between the type of argumentation, that is individual or collaborative argumentation, and the degree of formality ($\rho_{Q1,Q2} = 0.92, p = 0.376, n = 94$; $\rho_{Q1,Q6} = 0.79, p = 0.448, n = 95$; $\rho_{Q5,Q2} = -0.010, p = 0.926, n = 93$; $\rho_{Q5,Q6} =$

$0.191, p = 0.067, n = 93$).

However, the degree of formality seems to be connected to the overall benefit of computers in argumentation. Here, a higher amount of rules and axioms (Q5) correlates significantly with the usefulness of computers to improve classic face-to-face argumentation (Q12) ($\rho_{Q5,Q12} = 0.227, p = 0.031, n = 90$).

Research question 6: What are typical mistakes that occur in argumentation and do they differ in individual and collaborative argumentation settings?

To get further insights into the differences between individual and collaborative argumentation, the participants with expertise in teaching argumentation were asked what they see as the most common mistakes in different collaboration settings. Concerning the typical mistakes of individuals arguing, the resulting taxonomy of the open-ended question *"In your primary domain of interest, what are the most common mistakes made by students (or typical misconceptions) in formulating arguments on their own (that is, individually)?"* presented in Figure C.4 in the Appendix shows that there are two main problems: (1) the structure of arguments and (2) problems with evidences.

With respect to the first kind of error, that is the weak structure of arguments, typical problems include incompleteness of arguments, problems with ontology misconception as well as weak argument logic as exemplified by the following quotes:

> *"[Students provide] arguments that are formally incomplete [. . .]. This is the most common mistake. [. . .] arguments lacking on their logic, e.g., reasons that do not fully match the claim."*

> *"Perhaps the most common mistakes involve confusing critical issues, running together the questions of acceptable premises, relevance, and sufficiency or ground adequacy. Students might criticize premises for acceptability when the issue was whether the premises were relevant or sufficient."*

The second kind of error, that is problems with evidences, includes multiple facets such as the absence of evidences, the missing recognition of important evidences as well as problems with the distinction between claims and evidences. This is confirmed by the following representative quotes:

> *"[The students] did not really construct arguments they relied on opinions, they had difficulties developing one argument so they listed several possible arguments without real explanations."*

> *"Students sometimes repeat their standpoint as if it were self-evident instead of providing reasons to support it."*

> *"The students' arguments are sometimes weak because they do not provide supporting evidence as well."*

> *"There is sometimes a sort of difficulty to distinguish between standpoint/claim and argumentation."*

However, mistakes of individuals are only one side of the coin. The other side is mistakes that are connected to or even resulting from argumentation in group settings. Thus, the experts were also asked about typical problems that occur in collaborative argumentation. The resulting taxonomy of the open-ended question *"What kinds of problems occur most frequently when your students practice argumentation collaboratively with one another?"* are presented in Figure C.5 in the Appendix. Whereas a small set of experts just reported that the problems are the same as in individual argumentation, the majority of experts mentioned problems that are specific to collaborative argumentation. These problems included: First, (dis)agreement problems, that is students are either unable to elaborate different perspectives resulting in superficial agreement or even to recognize disagreement; they are biased towards their own point of view. Second, problems with the collaboration process such as not listing to others and unawareness of the work of others which leads to misunderstandings, repetitions and chaos. Finally, the compliance and agreement with typical rules of argumentation which is required to coordinate groups. These points are summed up in the following quotes from the experts:

> *"If group members agree for the most part on a shared position, [. . .] the tendency of simply compiling what everybody contributes without seeing the need to agree first on a common structure that makes sense. [. . .] If there is a controversy, there is lots of talk but less argument construction."*

> *"Students also lack sufficient skills in synthesizing and summarizing discussions, developing discussion threads, and learning how to move a discussion forward."*

> *"Unawareness (by the students) of all previous contributions, leading to repetitions of the same or related topics in different parts of the discussion map and rendering it disordered and less useful. Sometime this problem of awareness affects also the teachers, resulting in feedback of a lower quality than desired."*

> *"They do not listen to each other and attempt to counter views with which they disagree."*

Nevertheless, it is worth mentioning that some experts did not experience any problems with collaborative argumentation, especially if the collaboration process is structured as the following quotes show:

> *"The experience of students constructing arguments collaboratively is very positive."*

> *"Usually this goes more smoothly than one may expect."*

> *"However, I am selecting deliberately controversial issues in my class. First groups of 4 students who agree on a position construct and present their argument map, then conflicting groups are supposed to find either a compromise, a win-win solution, or a map that reflects the controversy. This approach leads to absolutely amazing results,*

> *simply the best you can achieve. Although the maps move to the background, this is really the point where you can observe creativity."*

Thus, one can summarize that collaboration is likely to extend the problems that may occur in individual argumentation settings. Nevertheless, these problems might be avoided by means of additional scaffolds that enable a fluent collaboration process.

4.2.5 Analysis & Feedback

Research question 7: What criteria are applied in practice to determine whether an argument is good?

An important step in the learning process is the provision of feedback to the learner. However, before concrete feedback can be given (either by a human tutor or an artificial intelligence agent) there must be a set of criteria which actually describes what is accepted as *good argument*. To collect these measures, the experts were asked the open-ended question *"In your primary domain of interest, what criteria do you apply to decide whether an argument is good?"*. As the taxonomy of answers presented in Figure C.1 in the Appendix shows, the most important point mentioned by the experts was the soundness / explicitness of the argument especially with respect to the structure. Additionally, the arguments must be seen in the context in which they occur. Typical quotes from the experts include:

> *"[...] the key to argumentation is the argumentative link, between a constellation of statements and thesis and notions, the topos, the warrant, that is where it is at."*

> *"I identify valid arguments by their conformance to a particular structure and also through consideration of the semantics of the argument."*

> *"The claims in an argument have to make sense given what is already known about the topic or domain."*

Whereas the conformance to provided guidelines (mostly in form of structure) is essential, the provision of facts and evidences for arguments is not less important. This way, it will get clear if the argument is well grounded as the following representative quote show:

> *"[...] whether an argument is "good" can be determined in this domain by its agreement with the facts"*

Nevertheless, the provision of hard facts in connection with a well-defined structure is only one side of the coin. On the other side, there are quite a couple of experts that focus on social norms as well as on the learning effect, that is an argumentation can be accepted as good once the arguers are able to learn from it in any way as the following quotes show:

"On the level of argumentation sequences, a rather long sequence of arguments, counter arguments and integrations may support learners to view problems from different perspectives."

"The argumentation is good if the argument stimulates learning, particularly higher-order (generative) learning."

"When teaching creativity it is the generation of distinct new perspectives within a discussion."

Directly connected to it is the last frequently mentioned point: The presence of rebuttals. Whenever alternative perspectives need to be discussed, it is required to promote an in-depth thinking about various explanations to gain insights into other people's views as the following quotes highlight:

"[...] I conceptualize the strength of an argument as its ability to withstand counterarguments. So a [...] criterion is the extent that arguments response to counterarguments (or plausible alternative explanations)."

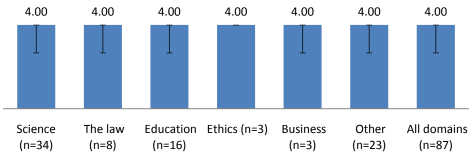
"[...] a good argument would involve cognitive conflict (a clash of beliefs) so that information in the discussion diversifies."

"I consider the persuasiveness of an argument in its ability to withstand and rebut objective criticisms."

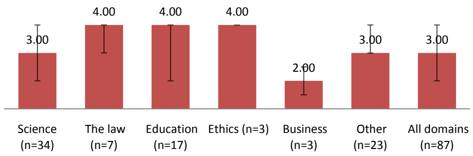
Once more, these various points-of-view highlight the complex nature of ill-defined problems such as argumentation. Overall, the experts focus more on the underlying principles of argumentation such as the structure and the social norms present in discussions than the concrete contents.

Research question 8: Is it possible to develop automated analysis features that can effectively analyze arguments and are there domain-specific differences?

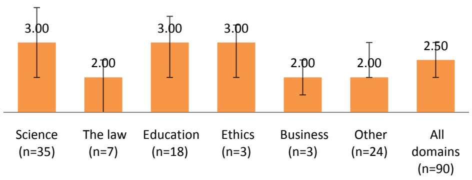
Having these criteria in mind, one can go on to the results shown in Figure 4.8. It shows that typical weaknesses and errors in argumentation can be identified, in the experts' opinions, by means of general and recurring patterns (see question 7) which is directly supported by the importance that the experts assigned to the structure as criteria. Here, especially those experts with a higher amount of teaching experience tend to believe more in patterns than those with less experience ($\rho_{T-Exp,Q7} = 0.223, p = 0.037, n = 92$), which is not surprising since teachers are expected to use reproducible patterns for assessments. However, in some domains such as science and business the experts think that these patterns are too complex to be detected automatically (see question 8). In addition, the experts are rather skeptical that general patterns alone are adequate to assess the overall quality of an argument (see question 9). The explanation therefore is obvious: Just analyzing the structure of an argument does



(a) Question 7: In my primary domain weaknesses and errors in arguments can be identified by general and recurring patterns.



(b) Question 8: In my domain of interest it would be possible to define a set of rules that could be used to automatically detect general and recurring patterns.



(c) Question 9: It is possible to assess the quality of an argument just by analyzing general patterns in the argument.

Figure 4.8: Evaluation of suitability of analysis and feedback methods for argumentation (Scale: 1=Strongly disagree, 2=Disagree, 3=Neither agree nor disagree, 4=Agree, 5=Strongly agree). Correlations: $\rho_{Q7,Q8} = 0.302, p = 0.005, n = 85$, $\rho_{Q7,Q9} = 0.087, p = 0.438, n = 82$, $\rho_{Q8,Q9} = 0.227, p = 0.039, n = 83$

not evaluate the other important criteria for good and educative arguments (such as social norms) mentioned above. Thus, the first part of the research question has to be negated: Whereas patterns are present in all domains, just an analysis of general patterns is not enough to judge the overall quality. Concerning the second part of the research question, that is the existence of domain-specific differences, a non-parametric Kruskal-Wallis test did not show any significant differences ($H_{Q7}(3) = 0.334, p = 0.953$; $H_{Q8}(3) = 1.121, p = 0.772$; $H_{Q9}(3) = 6.674, p = 0.083$).

Research question 9: Who is going to provide tutorial feedback in practical argumentation learning settings and when is it most effectively provided?

Based on the results of open-ended question *"What general types of feedback do you (or would you) give to students when they make oral or written arguments in your primary domain of interest?"* the taxonomy of answers shown in Figure C.6 in the Appendix indicates that one can distinguish three sources of feedback: instructor feedback, feedback from other group members (in collaborative settings) and automated feedback from artificial intelligence agents. Here, the favored approach of our experts is the instructor feedback. This is not surprising as most of the experts reported to be not familiar with artificial intelligence methods and — contradictory to an instructor — group work is not always present when teaching argumentation. In addition, the teachers are of course convinced of their typical role in education. However, when connecting these results to the kind of feedback it gets clear that teachers prefer to give feedback that is somewhat dependent on expertise in argumentation and, hence, cannot always be provided by student groups. In particular, teachers prefer to focus on structural feedback (e.g., asking for supporting evidences and highlight invalid parts or missing positions of the argument), prompts to reflect (e.g., asking questions to explore to motivation behind the argumentation process) as well as context specific feedback (e.g., correcting concrete arguments) as exemplified by the following quotes:

> *"I might use one or more of the models of argument to draw their attention to elements that are missing or poorly connected."*

> *"I explore the dimensions of their argument with them, pushing for alternative positions and reasons."*

> *"I try to help them see the abstract pattern of reasoning, in order to highlight the fallacy they committed."*

> *"I try to make students comment on the performance of themselves and of each other so that they themselves will be able to discover patterns of argumentation. For that purpose, I ask the students about (1) Their preparations and plans in relation to the actual process and outcome of argumentation. ("What happened and what had you expected?") (2) Analysis of the reasons and causes for (1). (Why did you foresee or plan that; What determined the actual process and outcome? (3) Evaluation. (What was really good in the argument and what could easily have been improved?) (4) Generalizations. (Would you prepare, plan and execute along the same lines if you had another chance?"*

> *"Most of the feedback would be in the forms of challenges e.g. how do you know?"*

Once the kind of feedback that should be provided is clear, an open question is still the time at which feedback should be given to the learners. In chapter 3.7 different feedback strategies and timings were described as part of the review. Even though there are multiple attempts present, it was not clear which one is preferred in practice.

The results of the survey highlight that the experts did not agree on a "best" time to provide feedback. Figure 4.9 shows that most experts (approx. 43.6%) prefer to provide feedback immediately after the error or problem occurred to ensure that the error context is still available, whereas another group of experts (approx. 35.9%) recommend to give feedback after the argument is over. The on demand approach (as present in LARGO, see chapter 3.9.7) is only rarely support (approx. 3.8%). The rest of the experts (approx. 16.7%) mainly reported that it depends on multiple factors such as the number and experience of participants or the general context.

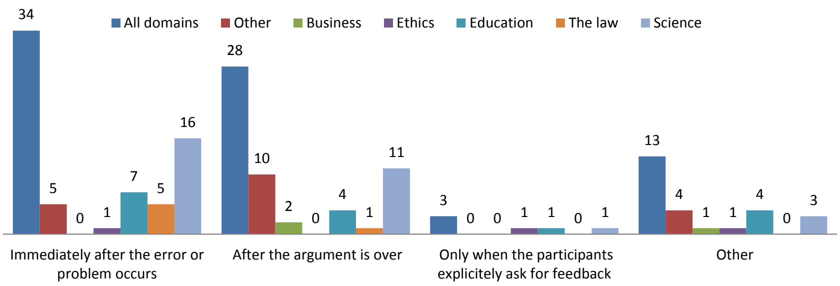


Figure 4.9: Preferred feedback strategies and timings. [Question 10: Feedback on errors and problems when engaging in argumentation learning is most effectively provided...]

Other observations

In addition to the concrete answers to the research questions, there were a couple of interesting correlations that might be considered when interpreting these results: First, the more knowledge of artificial intelligence technology experts have the more they believe that individual argumentation is appropriate for learning ($\rho_{AI,Q6} = 0.297, p = 0.004, n = 94$). Second, a higher amount of knowledge of artificial intelligence technology correlates with the believe that computer systems are able to support argumentation ($\rho_{AI,Q12} = 0.210, p = 0.046, n = 91$). Third and similar to the first point, the more experts believe that patterns can be automatically detected (Q8) and used to assess the quality of arguments (Q9), the more they believe computer systems can improve (Q12) or even replace (Q11) face-to-face argumentation ($\rho_{Q8,Q11} = 0.265, p = 0.014, n = 85$; $\rho_{Q8,Q12} = 0.284, p = 0.008, n = 85$; $\rho_{Q9,Q11} = 0.276, p = 0.009, n = 88$). Finally, there is a surprisingly negative correlation concerning the degree of formality (Q1) and the possibility to identify weaknesses by means of patterns (Q7) ($\rho_{Q1,Q7} = -0.261, p = 0.015, n = 87$).

4.2.6 Guidelines for Future Developments

Research question 10: What kind of flexibility is present in existing argumentation technology to face the challenges of various argumentation settings and domain-specific differences?

To get insights into the configuration and flexibility options of argumentation systems (existing one as well as those which are currently under development), the developers and system designers among the experts were asked the open-ended question *"Can you briefly describe the types of flexibility and/or configuration you provided in the design of your argumentation system (or systems)?"*. The resulting taxonomy of answers is shown in Figure C.8 in the Appendix. In general, the answers can be split up into three blocks.

The first block includes flexibility mechanisms that have been reported by about half of the experts. These mechanisms can be divided into the two categories (1) argument models and schemes and (2) user interface. As part of the former, that is the underlying argument model, typical configuration options here included (a) a flexible ontology in combination with a re-categorization of argument parts, (b) the definition of constraints such as scripts that require students to follow argument patterns (e.g., argument-counterargument chains) and (c) the assignment of weights to argument parts. Typical examples present in the quotes include:

> *"[. . .] the format of the debate, the ontology etc can be adapted to the circumstances."*

> *"[. . .] predefined rhetoric structure interaction grammar (sentence openers that can be used according to the previous act) manageable by the tutor (though previously to the use of the tool and not at run-time)"*

> *"We kept the constraints that the system imposes flexible: a discussion starting without any constraints and gradually increasing the rules as the discussion evolves."*

The latter, that is flexibility with respect to the user interface, contained (a) the ability to (dynamically) toggle system features on and off, (b) the definition and manipulation of visual shapes representing argument parts in addition with multiple different argument visualizations like graphs and texts, (c) the manual arrangement of arguments represented in a graphical fashion and (d) the configuration of fonts, colors and interface languages. This can be exemplified by means of the following quotes:

> *"Flexibility: the system offers (mainly for research purposes) a number of fading out mechanisms. For example, it is possible that the system support on how to construct an argument fades out after a certain time period, or times that the students have seen the supportive information, or based on a peer monitoring technique (that is students are asked to monitor the quality of their peers' [argumentation])"*

> *"Concerning argument visualization, users have the option of viewing the arguments of a debate in graphical or textual form, whichever suits their preference."*

> *"Finally, the collaborative features of the system can be toggled on and off to create a spectrum of interaction ranging from an individual learning environment to a collaborative work environment where student can discuss, critique, and share their work"*

The second block, which was mentioned by a fifth of the participants, contains analysis and feedback parts as well as the type of collaboration. As part of the former, that is the analysis and feedback configurability, the systems described by the experts can either support argumentation directly or indirectly by informing a tutor/moderator of problems:

> *"We have a fully automated proof search system in the background; if students have problems with a problem, then they can "appeal" to the Tutor. The Tutor passes the partial proof to the proof search system; the latter completes the partial proof and then*

> *the Tutor provides at first broad strategic advice to the student (based on the completed proof). If the students want to know more, they can continue the dialogue!"*

> *"The intelligent coaching is also flexible in a number of ways. Teachers can choose the level of depth to which the coach will offer assistance, how close a student argument must be to the experts in order to be accepted, etc."*

As part of the latter, that is the flexibility with respect to the collaboration, the assignment of roles and associated access rights, the definition of individual and collaborative work phases or the support of awareness mechanisms, e.g., an alert whenever a user modifies controlled parts of the argument was mentioned. The following quotes exemplify these points:

> *"[. . .] flexibility for communication (a-synchronous, synchronous, turn taking)"*

> *"Students use an interface to create an explanation. The explanations are then used as seed comments in an asynchronous discussion forum. 4 or 5 students are placed into a forum and instructed to reach a single best explanation."*

> *"Different kinds of awareness tools, designing alerts [. . .]"*

The third and last block contains features that have been reported rarely including technical flexibility, that is platform independence and flexible data representation, as well as the support for questions to provide guidance as done for instance in argumentation Vee diagrams (Nussbaum et al., 2007):

> *"The system is motivated by a number of questions which I call basic dialectical questions. Answers to different questions play functionally distinct roles, seeing them as answers to these questions helps make the structural differences intelligible."*

Nevertheless, a minority of system developers and designers (approx. 12.5%) reported that they did not implement any form of flexibility. In addition, the flexibility reported by the experts required in some cases programming skills that cannot be expected from the target groups (e.g., students and teachers) of such systems.

Research question 11: Are there any guidelines that should be complied for future developments based on the experts' experiences?

To benefit from the experiences collected during the development of existing argumentation systems, here a two-tier approach is followed. First, the design and development experts were asked the open-ended question *"In the argumentation system (or systems) that you have designed and/or developed, what would you say was the best feature? Why?"*. This way, there will be light thrown to the — in the experts' opinions — successful parts of past developments. Nevertheless, whereas the stories of success are likely to be caught, at least in parts, in the review (see chapter 3), this is rarely the case for failures in the past. Thus, the second part of this evaluation will focus on the lessons learned by means of answers to the open-ended question *"What* *would you say was the most important lesson (or lessons) learned in designing, developing and/or testing your argumentation system?"*. By combining the results of both questions, potential guidelines will be identified for future developments.

Concerning the *"best"* argumentation system features, the experts' answers have been categorized into five classes as shown in Figure C.7 in the Appendix: (1) an appropriate argument representation, (2) human-computer interaction aspects, (3) flexibility and configurability, (4) analysis and feedback as well as (5) pedagogical design.

Regarding the first of these five classes of responses, visual representation is seen as key to aiding argumentation, including multiple representations (e.g., graphs, tables, threads, etc.; see chapter 3.2) to adapt to situational needs:

> *"I have worked on developing a diagramming system for arguments. The best feature is the ability to represent graphically and perspicuously how the various components of the argument fit together, including dialectical components such as rebuttals or defeaters, and replies to these components."*

> *"The ability to (re-) present lines of argumentation in different formats, such as plain text, graphs, hierarchical structures and matrices."*

The second class, that is the human-computer interaction aspects highlighted as best feature, include multiple points: (1) functions to simplify the interaction with arguments (e.g., a minimal set of ontology elements, a search function to check if points had already been made), to simplify data collection for research (e.g., support for replays) or to support learning (e.g. provision of materials, self-explanation prompts), (2) the ease of use of the overall system (e.g. required training time, error avoidance by means of restrictions) as well as (3) the support for collaboration (e.g., supporting discussions among groups, role distribution, applying peer-reviews). Typical quotes that describe these features are:

> *"In general, flexibility and ease of use are essential. At hig[h]er levels of use, other features may become important, but I have not often seen that being the case."*

> *"By restricting students to posting and labeling their postings (and tagging each message label with team member ship, + or -) to a threaded discussion board, the students can easily see and monitor the flow of discussion within each argument thread using any typical threaded discussion board. Students can quickly identify exchanges between members from opposing teams, and respond appropriately to their opponents challenges/rebuttals."*

> *"Peer review, because students learn from each others' strengths and weaknesses."*

> *"[. . .] cut and paste from the chat and data sources into the diagram nodes — it helped bind the source materials and discussion content to the diagram content"*

> *"The best feature was to distribute roles and thereby changing expectations of what learners have to do, i.e. applying specific heuristics to construct sound arguments, and help them coordinate their interactive argumentation"*

Concerning the third class, that is flexibility, developers frequently reported a flexible argument model in their system as best feature. Through configurable ontologies, these systems are seen as able to scaffold the argument creation process. Further, the manipulation of arguments had been explicitly mentioned. Typical quotes include:

> *"Enabling collective argumentation with flexible ontologies."*

> *"The best feature is modifiability of the diagrams. Letting the students do that."*

With respect to the fourth class of best features (analysis and feedback), the answers were split into two parts: (a) automated feedback and (b) human feedback including moderators support. In contrast to the results presented in research questions 7 to 9 (see section 4.2.5), both methods balance each other. Examples are presented in the following quotes:

> *"Providing a feature that indicates topics that are the most "arguable" ones."*

> *"Enabling moderation in argumentation (moderation = non-intrusive caring)"*

The fifth and final class of features (pedagogical design) include the following specific features: (1) Support for group formation and the assignment of roles, (2) micro/macro-scripting in connection with fading of scaffolds as well as (3) game features.

> *"The best feature was to ask people to discuss topics that even experts in the domain are not willing to take a stand on. It is as if students have a sense of topics that are worth discussing. They appeal to everyday experience, etc, etc, instead of referring to teachers as authorities."*

> *"Roles and role switches (as learners do hardly ever switch roles)."*

> *"A fade out mechanism of the micro script that guides students' argumentation. If students are to internalize argumentation rules they have to understand that the system support will gradually fade out."*

> *"In our [. . .] tool the argumentation is represented as a sort of battle between two positions, advancing or retracting according to adding pro arguments or counter arguments (trying to reaching a central flag). This feature depicts the argumentative strengths of the positions and the complexity of the argumentative structure."*

As mentioned before, the stories of success are only one side of the coin. The other (not less important, but less frequently reported) side is the failures. According to the taxonomy presented in Figure C.9 in the Appendix, there is quite a large set of categories in which failures have been reported. However, most of the problems resulted from a rather complex design. Thus, *"keep it simple"* is perhaps the most important rule that, in the experts' opinions, should be followed. A well-elaborated argumentation model is useless if not fully understood and used correctly by the learners. As the responses suggest, this simplicity should apply both to the argument model underlying the system and the user interface. Furthermore, the "design for simplicity" rule influences the amount of training required: the less training needed to use the tool, the more motivated the learners will potentially be to use it. Motivation, as has been reported by Pinkwart et al. (2008) (see chapter 3.9.7), can be a key confounding factor in the research on argumentation systems, and surely is critical for the practical success of systems.

As a consequence, the provision of additional system features and functions (e.g., artificial intelligence support, scripts, etc.) should be carefully evaluated before using them, as they may increase the overall complexity which may lead to demotivation caused by a cognitive overload. This is confirmed in the following quotes:

> *"If there are easier (not necessarily better) alternatives (such as doing nothing) available, these will be preferred"*

> *"The need to "teach the teachers"; their minds are, sometimes, not fully open to the use of these systems and their appreciation at the outset of their potential benefits cannot be taken for granted. The result is that teachers seem to be more readily accepting a relatively simple visualization aid than a more developed one, or one that relies more massively on AI or similar. The time teachers dispose of for learning "revolutionary" systems is a-priori limited, and the burden to prove their worthwhileness is clearly on the developers' side."*

> *"Good design involves users and understanding why they do what they do. Argumentation fares best in contexts where it is commonly used and people understand what uses it has."*

> *"Not because AI [artificial intelligence] could be used [. . .] it should be used."*

> *"My most important lesson was that it is possible to design a system that with the right instruction is so compelling for people, that they use it consistently for performing their discussion. And that the system is even so compelling that people use it hypercorrectly, namely they use it to re-do their discussion after they had erroneously first performed the discussion in a chat; they feel they have to 'fill-in' the system by providing their input."*

In addition, some considerations from the more general fields of software engineering and human-computer interaction, related to piloting and user studies, have also been found to be valid for the specific area of argumentation systems:

> *"Do not leave the lab and go to the classroom without being very sure that the system will work there."*

> *"Evaluation with non-expert end users is very important."*

Another lesson learned is that empirical evaluation may not always result in direct success, even though an approach may be promising. Due to the complex nature of argumentation, experience shows that it requires a large amount of training for students to actually improve in argumentation. Thus, another guideline might be: Do not get demotivated too fast. This is supported by the following quotes:

> *"Another lesson learned is that it is hard to demonstrate with pre-post exams over short time periods the type of deep learning that most of us hope to create in students, so we need to develop other metrics and approaches to show that these system are helpful for more than just memorizing facts (for which a worksheet is a perfectly fine teaching tool)."*

> *"That the positive effects of the script and the prompts I used did not sustain - almost immediately after the script was taken away, learners fell back on their own argumentation strategies that they had before already. When the script was there again, they however again were better able to produce more high-quality argumentation."*

As a final set of "lessons learned", some respondents noting that a graphical representation of argumentative structures per se will not lead to improved argumentation skills. It is important to focus on the strengths of the representation and to be aware of the weaknesses as already reported in research question 3. Even though graph structures have been established as standard for argumentation systems, they are not always the best solution. In addition, there are other factors and circumstances (e.g. number of participants, gender, etc.) that may require a more in-depth analysis.

> *"That graphical ("boxes and arrow") representation of arguments is not always helpful, especially for people not used to such representations (e.g. police investigators and lawyers)."*

> *"I believe that we have not paid enough attention to the design of an argument diagram. The perceptual feature of a diagram is what is supposed to make them better than for example an argumentative text. However, I do not mean that a computer tool should provide this design by itself, it is very well possible that it is the student who needs to pay attention to the design of his/ her own diagram in order to understand or communicate about it."*

> *"[. . .] graphical objects may be exciting, but the power is in encapsulation of an idea, and the ability to position, juxtapose, flow, oppose, connect, etc. it's not the graph that is the key, but the moveability (mobility?) and flexibility of text that are powerful"*

> *"For example, the number and type of responses posted in reply to a stated argument can be affected by who (male/female, active/reflective learning style), when (day of week, response time), how (conversational/expository style) one expresses the argument"*

This supports the results of the empirical studies presented in the review (see chapter 3.9).

4.3 Discussion

The results of the survey show that at least some of the challenges, which have been identified in the review in chapter 3.10, have been recognized by the argumentation community.

On the one hand, the role of collaboration has become more important. Whereas many existing argumentation tools are designed to support individual argumentation, the trend clearly points to support for collaboration. Here, the experts agree that collaboration is the key to argumentation learning as it involves different points-of-views which are essential for active discourse. However, the support of collaboration is complex, because collaborative argumentation is likely to cause additional problems compared to individual argumentation which have to be dealt with. Thus, the computer is required to provide additional scaffolds which, on the one hand, supports the collaboration process but, on the other hand, must not cause cognitive overload. This balancing act is one of the ongoing challenges. However, collaboration is only one part of a successful argumentation process. Instead, the process involves multiple phases and tasks that involve individual work as well. Thus, a promising scripting approach should account for the complete process. A collection of phases and tasks as presented in Figure 4.3 can serve as starting point for future research in this area. This way, yet unsupported parts of the reference model can be evaluated more systematically in order to find ways to further improve the effectiveness of scripting approaches with respect to learning.

On the other hand, the incorporation of flexibility mechanisms got more into focus. As concluded in the review, the concrete effect of single factors such as visualizations and ontologies is still unclear. Whereas the experts are convinced that visualizations are helpful in general and in particular for learning purposes, the potential downsides have been recognized as well. In connection with a pre-defined and often times fixed ontology, the effects cannot be evaluated in an adequate way. Instead, one would need multiple tools which would implicitly affect additional factors (e.g., the interface) that are likely to have influence on the results as well. Thus, ongoing developments are intended to be configurable with respect to the interface or the underlying argumentation model.

In addition to the results of the review, the survey revealed further insights into other aspects such as the role of the computer which is considered as supporting medium instead of a replacement. This understanding implies a set of new challenges for all parties involved in the education of argumentation skills. While the computer is accepted as adequate tool to create a shared understanding (van Gelder, 2003) and to enable communication between (even spatially separated) people, there are still lacks concerning the analytical capabilities. Thus, the workload of teachers will stay high, because pure structural analyses methods which have been applied to argumentation in the past (see chapters 3.7 and 3.9.7) are not enough to provide adequate feedback. Instead, additional criteria such as social norms and the context in which argumentation takes place should be involved in the analysis which directly implicates new challenges for researchers and developers. A glimmer of hope, however, is that the concrete content, which is unequally harder to assess in ill-defined domains such as argumentation (see chapter 2.3), do not seem to play the central role. Therefore, existing methods of artificial intelligence may be adapted to fit the needs of teachers as well as of learners. However, one could also argue that the focus on the structural level has evolved from necessity, that is time limitations and the missing of objective rules.

The perhaps most important results of the survey are, however, the lessons learned. As mentioned before, a literature review usually only reflects the stories of success. Nevertheless, it is essential to learn from mistakes and to avoid them in future. Thus, one should always keep in mind the golden rule: *"Keep it simple"*.

At first glance, this seems to be completely contradictory to the points mentioned above. How should there be scaffolds for argument phases and tasks in addition with artificial intelligence support to provide learners with tutorial feedback while — at the same time — avoiding any unnecessary complexity? The answer is simple: Use it when it is appropriate. The resulting challenge, however, is to determine when, that is in which phase and which task, which means (e.g., type of visualization or feedback) is actually appropriate. To do so, a flexible argumentation framework is required that is capable of easily manipulating single variables to enable systematic research on argumentation. This supports the conclusion drawn from the review.

Together with the results of the review, this survey provided answers to research questions one and two (cf. chapter 1.2). Based on these answers the next chapter will summarize requirements on a generic framework that can be used to investigate most of the open challenges in the area of computer-supported argumentation. After that, chapters 6 and 7 will provide an answer to research question three.

4.4 Limitations

Even though these results are promising, the interpretation must be done with care, because they are limited in several ways. First, many of the experts who participated in the survey were mainly from the area of computer-supported collaborative learning. As such, it is not surprising that these experts would highlight the importance of collaboration for learning. In addition, some of the argumentation domains (e.g., ethics) were underrepresented which may have influenced the results. Furthermore, one should be aware of the point that the experts' answers are, by definition, subjective: it was asked for their personal points-of-view. Yet, one can argue that the aggregated opinions of experts have a value. Finally, some of the question formulations do imply collaborative and spatially separated settings, which also may have influenced the answers of the participating experts.

5 Requirements of a Generic Argumentation Framework

The review of existing argumentation systems and approaches (see chapter 3) in combination with the survey among argumentation experts (see chapter 4) identified, on the one hand, best practices applied in existing argumentation systems. On the other hand, however, it revealed a couple of open challenges which highlighted the need for a generic argumentation framework. The main goal of a generic argumentation framework is twofold. First, it should cover and integrate best practices identified in the review (chapter 3) and the survey among experts (chapter 4). Second, it should be applicable to solve open problems and resulting challenges. As a direct consequence, research could be done in a more systematic way. Currently, a test of the influence of a single variable would require (in most cases) multiple tools that differ in this variable. However, the use of multiple tools would also affect other variables and, thus, is likely to have impact on the outcomes. Here, an example is different user interfaces which are likely to influence the resulting arguments. Imaginable reasons for that may be either a more complicated or just a different way to create arguments. In contrast, a generic framework that allows the modification of single factors easily via configuration mechanisms would promote more comparable results in the area of argumentation research. In teaching, the flexibility of such a generic framework would be beneficial for their acceptance. Instead of adjusting the teaching to the tool's restrictions, the tool would be capable to deal with changing requirements.

In this chapter, the requirements that must be met by a highly flexible tool will be collected. Therefore, the preliminary results presented in (Loll et al., 2011b) will be picked up and extended. This list of requirements will use the typical split into *non-functional* and *functional* requirements. For each of these requirements a description as well as a set of concrete criteria to evaluate the fulfillment of the requirements will be provided. The criteria will be consecutively numbered. By means of references to other chapters, the motivation behind the single requirements will be highlighted.

5.1 Non-Functional Requirements

A non-functional requirement (NFR) can be understood as "*an attribute of or a constraint on a system*" (Glinz, 2007, p. 25).[1] The fulfillment of non-functional re-

[1] Other definitions can be found in (Chung & do Prado Leite, 2009)

quirements is a precondition for the successful integration of concrete functionalities, because non-functional requirements *"play a critical role in the area of architectural design"* (Chung & do Prado Leite, 2009, p. 369).

To satisfy the requirements on a generic argumentation framework that is capable to promote structured research on argumentation the following non-functional requirements are central.

5.1.1 Flexibility & Configurability (NFR1)

The framework must be capable to adapt to situation dependent needs. By means of a modular approach, the functionalities present in the framework can be combined to support real-world scenarios in different application contexts. In research, for instance, there could be two combinations of functionalities which differ only in a single point in order to evaluate the influence of it on the outcomes. Thus, this single point would be used as independent variable in a research setting. However, in other contexts such as education, various combinations of functionalities can be used to provide support for different learning scenarios such as brainstorming, group discussions and individual conclusion drawing. The configuration of these combinations should be done in a central place.

Motivation, sources & examples Chapters 3.9, 3.10 & 4.2.6

Criterion 1 The functionalities of the system must be modular.

Criterion 2 The configuration of functionality combinations should be done in a central place.

5.1.2 Extensibility (NFR2)

Research is an area which is highly dependent on new ideas. Therefore, the functions that are provided by a general framework are likely to be not sufficient for certain research projects. Thus, the framework should be easily extendable with new functions. In addition, external software should be able to communicate with the framework. This way, it would be possible to reuse existing work in the context of this framework.

Motivation, sources & examples Chapters 3.8 & 4.2.6; required to provide generality

Criterion 3 The modular set of functionalities should be easily extendable on the code level. This should be achieved by means of a common interface.

Criterion 4 The framework should be able to communicate with external applications.

5.1.3 Performance & Scalability (NFR3)

The system's responsiveness must be high enough to not be perceived as disturbing. In addition, it must scale up so that typical classroom sizes of users (approx. 30) are supported in order to be applicable to typical research and education contexts.

Motivation, sources & examples Chapters 3.5 & 4.2.4

Criterion 5 There should be no noteworthy lag between frequently executed users' actions and the system's response (that is, usually less than 1 second).

Criterion 6 The framework must scale up in an adequate way (with respect to performance) to support typical classroom sizes of users (approx. 30).

5.1.4 Loose Coupling (NFR4)

Directly connected to the performance, is the loose coupling of system parts. Here, it is important to allow the distribution of workload to avoid expensive high-end machines. In addition, a distribution of tasks (e.g., presentation of graphical argument representation, calculation of feedback by means of artificial intelligence, etc.) is also beneficial for the development phase of a system, because the components can be implemented independent from each other. Thus, a well-defined communication interface is essential to enable communication between system components and to allow for a spatial distribution.

Motivation, sources & examples Chapter 3.8; general software engineering concepts

Criterion 7 The framework should make use of modular concepts in order to allow a spatial distribution of system components. These components should communicate via a well-defined interface.

5.1.5 Usability & Motivation (NFR5)

Field studies are a typical application area of research tools. Here, motivation plays a key role for success of tools. To avoid a demotivation of participants, the interface's usability is an important factor. A similar picture is drawn when using the framework in schools or universities for education. Here, not only the students are inexperienced with the tool, but also the tutor is in most cases, which makes it unequally harder.

Nevertheless the interface's usability is only one side of the coin. The other one is technological usability. This implicates that the system does only require minimal preparations. That is, on the client side, any installation routines or firewall configurations which are typical sources of frustration in scholar settings when using networked — or more general computer — technology should be avoided. On the server's side, the amount of required preparations should be as low as possible.

Motivation, sources & examples Chapters 3.9.7, 3.10 & 4.2.6; dialogue with researchers and teachers

Criterion 8 The user interface of the system should be intuitive, that is, the amount of required training should be minimal (not more than 15 minutes).

Criterion 9 No installation should be required on the client side.

Criterion 10 The installation requirements on the server side should be minimal.

5.1.6 Openness (NFR6)

One of the main purposes of a general framework is to promote more comparable results. Therefrom, the framework should only use technology which can be distributed freely without paying expensive licenses. To benefit from other ideas and developments, the source code should be modifiable by other system designers as well. Thus, the wheel has not to be reinvented every time. In addition to the technological openness, additional concepts such as internationalization fell in this category.

Motivation, sources & examples Chapters 3.8 3.10; basic research concepts

Criterion 11 External libraries and technologies used in the framework as well as the framework itself should be open-source or at least freely available.

Criterion 12 Internationalization should be supported by the framework.

5.1.7 Platform Independence (NFR7)

Even though today most end-user computers are running Windows, this cannot be assumed for all settings. Further, other platforms than personal computers, for instance tablets, are imaginable in future as well. Thus, the framework should be capable to work across different platform settings, allowing them to work together.

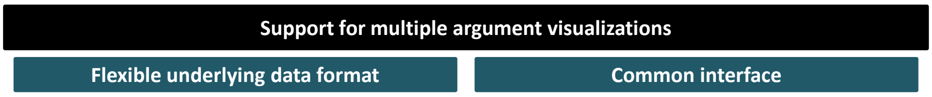
Motivation, sources & examples Dialogue with teachers; practical experience

Criterion 13 The framework should work platform independent.

5.2 Functional Requirements

Once the non-functional requirements are clear there is a set of functional requirements (FR) that need to be considered. Whereas a list of functions can be of nearly unlimited length, here the focus will be set on the requirements that can be directly derived from the review and the survey. The functional requirements are summarized by high-level topics (presented in teal boxes). However, each topic will get decomposed into multiple sub-topics (presented in violet boxes). These (sub-)topics are not mutually exclusive related to the functional requirement. In fact, they may occur in multiple functional requirements. The lists of criteria will reflect requirements that occur in multiple areas only report once. In some cases, examples (presented in red boxes) will be provided. Nevertheless, they are (in contrast to the (sub-)topics) not intended to serve as complete list, but for illustrative purposes.

Functional Requirement ID - 1 Support for multiple argument visualizations



Motivation, sources & examples Chapters 3.2, 3.9.2 & 3.10

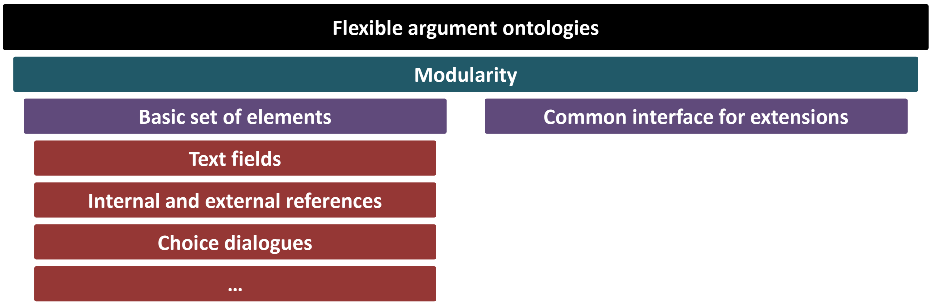
Description The framework should be capable to support different visualizations for arguments, e.g., graphs, tables or threads. In order to allow the exchange of arguments created in these visualizations as well as to allow the simultaneous use of multiple views, the underlying data format should be flexible, that is, independent from the concrete visualization. By means of a generic interface, the exchange of this flexible format will be simplified further.

Criterion 14 The framework should support multiple visualizations.

Criterion 15 The data format should be flexible to enable visualization independent data exchange as well as future extensions.

Criterion 16 The system should provide a common interface for the data exchange.

Functional Requirement ID - 2 Flexible argument ontologies



Motivation, sources & examples Chapter 3.6, 3.10 & 4.2.6

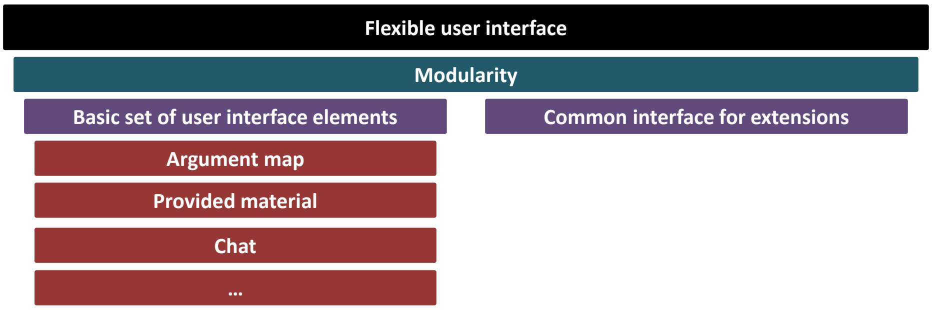
Description To enable the framework to deal with domain-specific argumentation differences, it should provide support to use different ontologies, that is, the underlying argumentation model must be configurable. These ontologies should be flexible so that they can be adapted to situational needs.

Thus, there should be a basic set of commonly used elements. These elements include the possibility for users to provide a text to actually argue, to link to existing resources such as given texts or external web pages and to assign scores. This set of elements should be extendable. In order to keep the development costs low, there should be a shared interface for all elements that can be used to define the underlying argument model.

Criterion 17 The underlying argument model should be modularly configurable.

Criterion 18 The framework should provide a set of elements that are typically used in argumentation systems (text fields, internal and external references, numeric assessments, choice dialogues).

Functional Requirement ID - 3 Flexible user interface



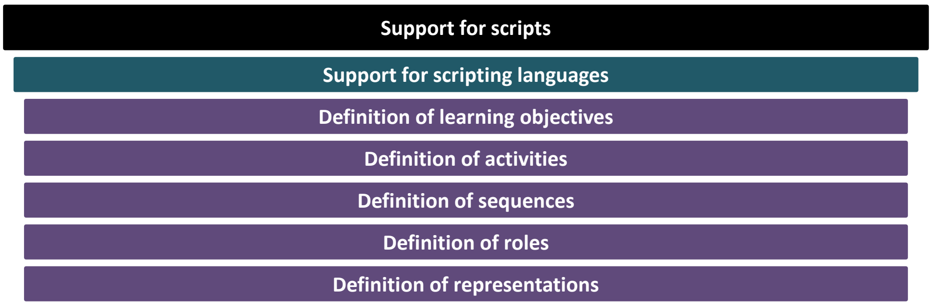
Motivation, sources & examples Chapters 3.10 & 4.2.6

Description Whereas flexibility on the ontology level is essential, another important aspect is the flexibility of the user interface. Typical requirements here include the provision of additional means for collaborative settings in order to increase the overall awareness of actions. Examples are a list of active users or a chat for communication. However, different collaboration settings are only one side of the coin. The other one is independent from the number of users. Here, examples include the provision of a text containing background information or additional panels that show node or relation details that are hidden from the argument map in order to increase the overview.

Criterion 19 The graphical user interface should be modularly configurable.

Criterion 20 The framework should provide a set of user interface parts that are typically used in argumentation systems.

Functional Requirement ID - 4 Support for scripts

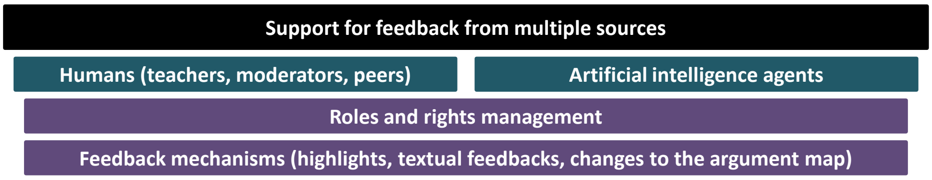


Motivation, sources & examples Chapter 3.5 & 4.2.3

Description The framework should be able to support various activities in the overall argumentation process. Therefore, the typical underlying concepts of a collaboration and learning script should be configurable. These configurations include the definition of common components present in scripts, that is, the definition of learning objectives, activities, sequences, roles (in connection with role specific rights) and representations. By combining definition packages, complete learning activities should be supportable.

Criterion 21 The framework should have a scripting engine that allows the definition and execution of collaboration and learning scripts.

Functional Requirement ID - 5 Support for feedback from multiple sources



Motivation, sources & examples Chapters 3.7, 3.9.7, 4.2.5 & 4.3

Description The framework is intended to be used in learning scenarios. Here, the provision of feedback is essential to recognize mistakes. However, feedback can be provided from various sources including teachers as well as artificial intelligence agents. Thus, the framework should enable all kinds of sources to actually provide feedback to the learner. Further, these sources should be able to highlight or even modify the argument, that is, they should not be artificially limited in their capabilities to interact with the system. Instead, the specification of roles and rights should be used to define the capabilities of, for instance, a feedback engine. This way, even a complete simulation of a student is imaginable.

Criterion 22 The framework should have a roles and rights management.

Criterion 23 The provision of feedback from multiple sources (internal and external) should be supported.

Functional Requirement ID - 6 Logging and persistence mechanisms

Motivation, sources & examples Chapter 3.8

Description All argumentation moves should be logged in detail. This way, it should be possible to reconstruct the whole argumentation process. This is essential for multiple reasons: On the one hand, teachers are able to get the context in which a mistake evolves and researchers may be able to recognize potential points of interest. On the other hand, a detailed action log would enable the possibility to export the resulting argument in foreign formats used in other tools. Based on these action-based logging mechanisms, a snapshot of the current state can always be created to exchange light-weighted files.

Criterion 24 It should be possible to watch a replay of the argumentation process based on action-based logs.

Criterion 25 The framework should support the creation of snapshots in form of state-based logs.

Criterion 26 There should be an export/import function to exchange arguments.

Functional Requirement ID - 7 Support for individual and collaborative use

Motivation, sources & examples Chapters 3.5, 4.2.3, 4.2.4 & 4.3

Description The framework should be capable to support individual argumentation as well as (a)synchronous collaboration in typical group sizes (up to 30 people) present in scholar settings. Along with the support for collaboration comes the need for adequate concurrency control, awareness and communication mechanisms. This means in particular that users should be aware of what other users are doing and they should be able to communicate with each other without the need for external applications. Potential conflicts that are likely to occur in synchronous collaboration settings, e.g., modification of argument parts at the same time, must be handled in an adequate way to avoid a data loss which would cause demotivation among arguers.

Criterion 27 The framework should provide communication facilities to support arguing in groups.

Criterion 28 The framework should provide awareness aids to the users to promote a fluent collaboration (meta-data who created which part of the argument and when; who is currently working on which part).

Criterion 29 In collaborative settings, there has to be a concurrency control that ensures that a data loss is avoided.

6 The LASAD Framework - Architecture

Based on the requirements collected in chapter 5 the fundament of a generic argumentation framework called *LASAD (Learning to Argue: Generalized Support Across Domains)* will be presented in this chapter before a reference implementation of this architecture as proof of concept will be described in chapter 7. In order to connect the identified requirements, there will be references to the non-functional requirements (NFRs) and functional requirements (FR) provided in both this chapter and the next one. At the end of chapter 7, there will be a table which summarizes how the requirements were fulfilled.

6.1 Architecture

As shown in Figure 6.1, the proposed framework is using a classic layered architecture. It consists of three layers: the client layer, the server layer, and the data layer. Each layer is only able to communicate with its direct neighbor layer via a well-defined interface. This way, it is possible to use and modify each layer independent from each other. In addition, each layer can be distributed to another machine, which is beneficial for load distribution and, hence, performance (→ *NFR3: Performance & Scalability*).

6.1.1 The Client Layer: Visualization & Interaction

The top level, that is the client layer, is the 'window' to the system. Here, users are able to argue via a graphical user interface that allows them to create and modify graphical representations of arguments as well as to communicate with others (in collaborative settings). Each client is responsible for the representation of the argument structures itself. That is, one client could use a graph-based representation whereas another one is using, for instance, a table-based visualization or even both (→ *FR1: Support for multiple argument visualizations*). However, the client layer is not restricted to human users, but artificial intelligence (AI) clients can also be added (→ *FR5: Support for feedback from multiple sources*). The latter may be used to pinpoint the user to possible weaknesses in their argument structure based on analyses of the argument (an overview of feedback techniques for argumentation support is given in Scheuer

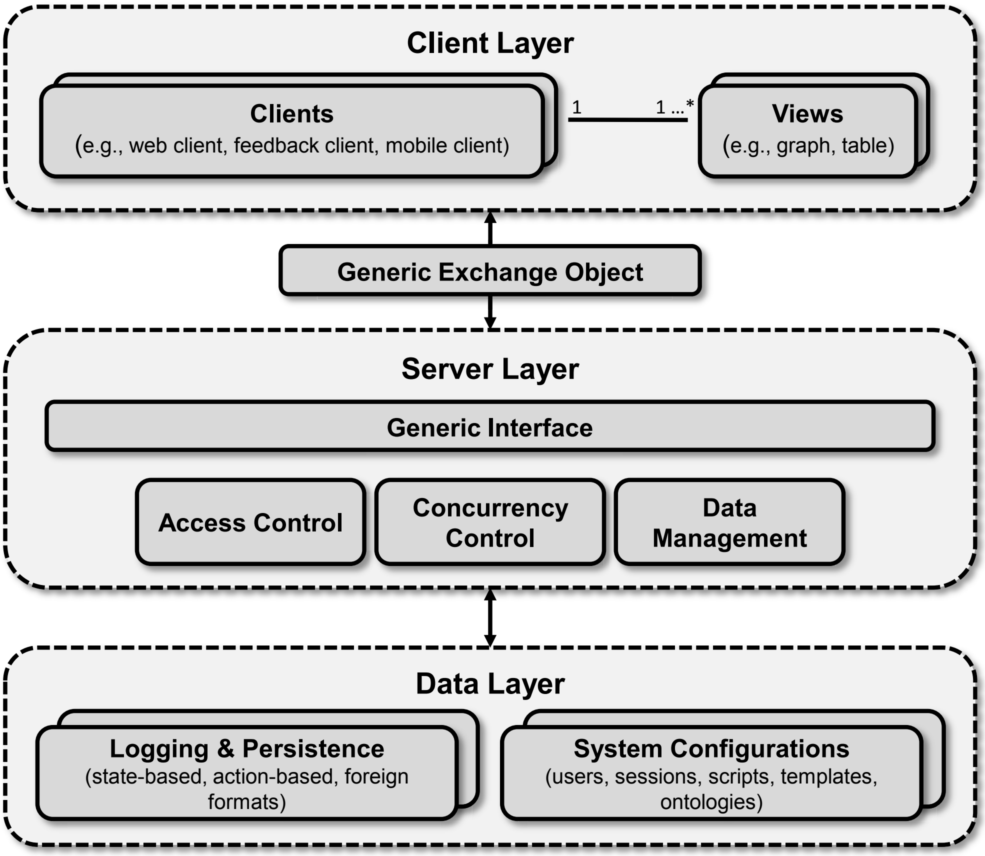


Figure 6.1: Architecture of the LASAD framework

et al. (2010, 2011)). To guarantee flexibility, each client is technically able to do the same actions, that is, even an AI client may modify the argument if desired. The different clients may not only represent different roles (such as human user or artificial intelligence feedback agent) with assigned rights, but can also provide different views (e.g., a table-based view instead of the graph-based view as done by Zhao (2011)) (see chapter 9.1.2) on the same logical argument provided by means of a generic data structure (see below). This way, flexibility on the representation and visualization level is achieved. In order to communicate with the server, there will be simple, but generic objects. The structure of these objects is shown in Figure 6.2.

Here, each client action, for instance a modification of the argument representation or the sending of a chat message, is translated into an *action* object. In order to allow filtering of relevant actions on the server side, each action is identified by means of a *category* (e.g., *"MAP"*) and a *command* (e.g., *"DELETE-ELEMENT"*). To specify the change that has been made to the argument representation, there is a generic

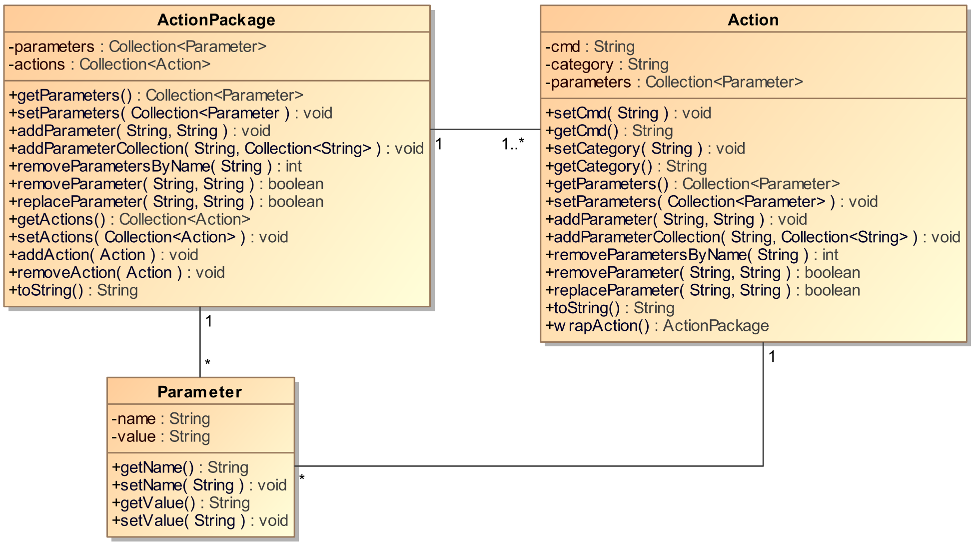


Figure 6.2: Structure of communication objects

set of *parameters* for each action. Each parameter consists of a pair of two parts, *name* and *value*. An example of a typical parameter is *"TEXT"* (*name*) and *"In my opinion, argumentation representations are essential to discuss about important issues"* (*value*). In order to reduce the network load, one or multiple actions can be collected in an *action package*. These action packages, however, can be enriched with additional parameters, for instance to identify the client. Via a well-defined interface, these generic objects will be exchanged between clients and server. This way, the client's implementation is independent from the server's one (→ *NFR4: Loose Coupling*). Thus, the programming language can be freely chosen (→ *NFR7: Platform Independence*).

In addition to the client's role to support users in creating argument structures, additional clients can be used as extension point to the overall framework (→ *NFR: Extensibility*). Imaginable are, for instance, clients that are not designed to modify the argument structure, but that log the actions in a different way than the server or that act as simulated users.

6.1.2 The Server Layer: Management

On the second layer, all incoming requests and actions from the connected clients are managed, that is requests will be answered based on the information stored in the data layer, and actions will be distributed to other clients that are working on the same argument. Therefore, the server's interface is required to be accessible from various clients that may be implemented in different programming languages, which is the server's central contribution to the overall flexibility of the system (→ *NFR7: Platform Independence*). Additional tasks that the server is expected to handle are the concurrency and access control (→ *FR7: Support for individual & collaborative use*). Here, the server has to sort all incoming actions so that the number of potential conflicts caused by concurrent client access on the same object is minimized. By means of additional information from the database, the server is also required to check whether a client is allowed to execute certain actions such as element modifications. Overall, the server acts as control instance between client and database.

6.1.3 The Data Layer: Storage & Flexibility

The data layer is the key to flexibility of the framework. By means of XML configuration files, it is possible to tailor the system to domain or application specific needs. These configurations include (a) the definition of user accounts, (b) the definition of the underlying argument model, that is the ontology (→ *FR2: Flexible argumentation ontologies*), (c) the definition of user interface components that are required to be shown in a user client (→ *FR3: Flexible user interface*) as well as the argumentation context, and (d) the definition of argumentation sessions that are based on a combination of an argument model and an user interface definition. Together these definitions can be used to define a complete argumentation process in form of a collaboration and/or learning script (→ *FR4: Support for scripts*). These configuration mechanisms will be discussed in more detail in the next section.

In addition to the configuration of the system, the data layer provides support for detailed action logging (→ *FR6: Logging and persistence mechanisms*). Here, a two-level approach consisting of action-based logging as well as state-based logging is used. The underlying concepts of these logging mechanisms as well as the motivation behind them will be explained further in section 6.3.

6.2 Configuration Mechanisms

As mentioned in the previous section, the configuration of the LASAD framework consists of multiple parts. An overview of these parts is shown in Figure 6.3. In this section, each part will be described in detail.

6.2.1 Definition of Users

The first part of the configuration is the definition of system users. A user definition is described by three components: a nickname, a password and a role. Each role is connected to a set of rights. Possible rights would, for instance, include the ability to create or delete argument parts as well as communication with others and the

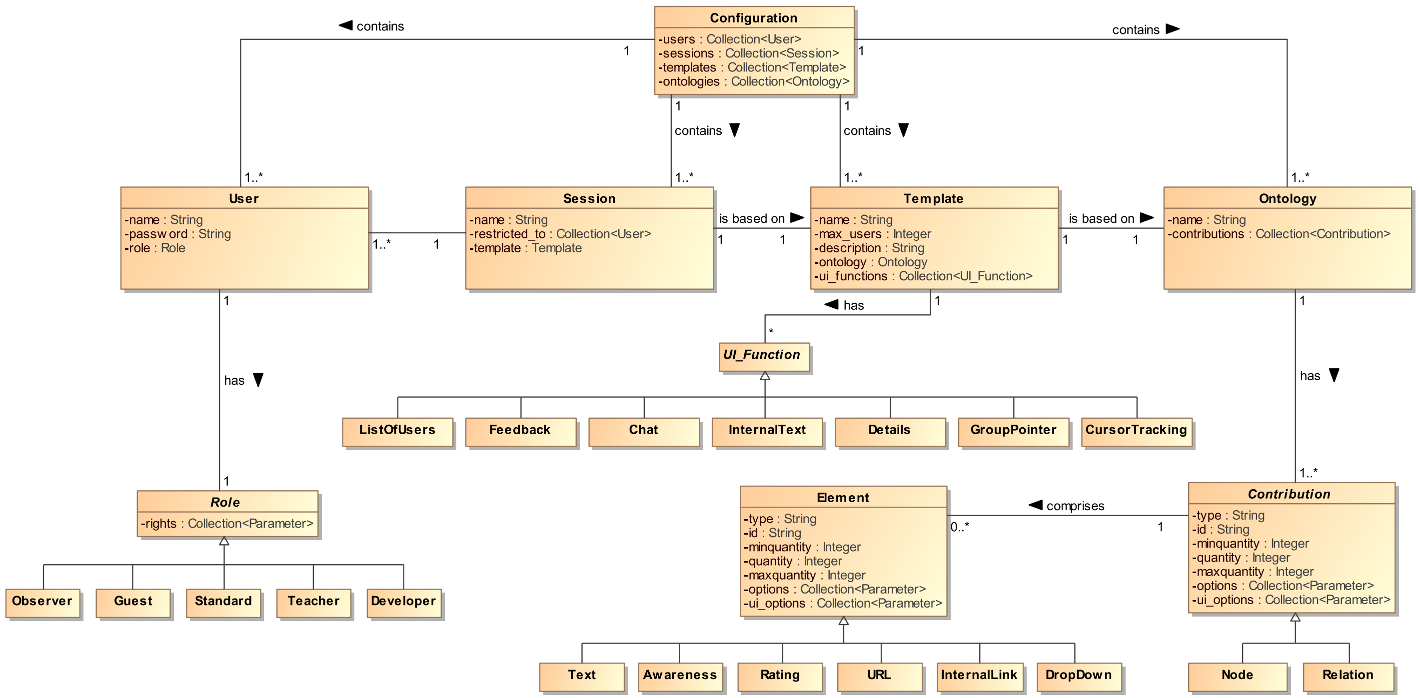


Figure 6.3: Configuration of LASAD

highlighting of elements. Typical roles that are likely to be used are, for example, teacher, student or moderator (other examples can be found in chapter 3.5).

6.2.2 Definition of the Underlying Argument Model

The second part of the configuration is the definition of the underlying argument model, that is, the ontology. It can be understood as a definition of argument structures. These structures can be divided into top-level elements and sub-level elements. To illustrate: In a graph-based visualization, the nodes and relations are the top-level elements. These, in turn, comprise one or more sub-level elements such as text fields, references, etc. A set of typically used sub-level elements in existing tools (based on the review presented in chapter 3) is given in Table 6.1.

Sub-level element	**Description**
Text	A text field (a single line) or text area (multiple lines) to add the concrete content to an element of the argument map.
Awareness	This sub-level element is used to provide information about the author of an element including the username and the creation date.
Internal link	Allows to create a relation between a prescribed text and the element.
URL	A reference to an external web page.
Rating	Assigns a score to the element. The score can be used, for instance, to assess the quality of the element's content. In addition to the pure number, the element has the possibility to provide a label for it.
Drop-down	A set of given alternatives a user can choose from.

Table 6.1: Typical sub-level elements used in argument models

To provide flexibility with respect to the underlying argument model, there must be configurations for each element. Each element is described by a *unique identifier* (*elementid*; e.g., fact), a description of the *element type* (e.g., text field) as well as a set of element and visualization (or UI) options. In addition, each element's appearance can be limited. This way, it is possible to define how many instances of, for instance, a text field are allowed to be added to a top-level element during runtime. However, the element options define the settings that are specific for this element. An example is a standard text provided in a text field or the score range of a believability assessment element. In addition, the visualization options provide additional information how to display the element. Even though the client is expected to draw elements in an adequate way, the provision of additional parameters may be beneficial for understanding. Here, typical examples are the kind of border used for a node in a graph-based environment or the background color of a table-cell in a table-based visualization.

In order to create these element definitions, XML is used. XML is, on the one hand, human readable and, on the other hand, platform independent. In Appendix G, two complete example configurations are given. Here, the following short example will clarify how an element in LASAD is configured:

```
<element elementid="content" elementtype="text" quantity="1"  minquantity
    ="1" maxquantity="1">
  <elementoptions />
  <uisettings background-color="#FFFFFF" font-color="#000000" minheight="
      28" />
</element>
```

In this example, a sub-level element of type *"text"* is described. It is limited to be created only once in the top-level element. If the *maxquantity* variable would be set to 2, there could be another element of the same type added during runtime. Since there are no *elementoptions* specified, the text field will appear empty without a given standard text in it. The optional *uisettings* define that the background of the text field should be in white (*#FFFFFF*) and the color of the text that can be entered should be in black (*#000000*) color. Further, the minimal height of the element is restricted to 28 pixels. The complete underlying argument model can be specified this way.

6.2.3 Definition of User Interface Components and the Argumentation Context

The definition of the underlying argument model is only one side of the coin. The other one is the available user interface elements and the context in which argumentation takes place. This is the third part of the configuration. The former includes typical elements that are used in existing argumentation systems. Examples include the provision of a prescribed text or, in collaborative settings, the provision of a chat. In addition, the argumentation context is specified. This context contains, for instance, the definition whether the argumentation should take place individually or collaboratively and how many users are allowed to argue at a time. In the framework, this definition is described as argumentation template. A list of typical user interface elements and context settings which are used in existing argumentation systems is given in Table 6.2.

Option	**Description**
Chat	Enable or disable a simple text chat among users.
Cursor tracking	Enable or disable the submission of cursor positions for additional awareness to all other users who are actively participating at the argumentation.
Group pointer	Enable or disable a shared indicator that can be used to point to a specific part of the argument map
Element's details	Enable or disable an additional window for each element on the argumentation map (i.e., nodes and relations) to show possibly hidden sub-level elements such as additional notes that should not appear on the argumentation map directly.
Feedback	Enable the support to show feedback from teachers or artificial intelligence clients.

Option	Description
List of users	Enable or disable a list of users that are actively working on an argument map.
Number of users	Define the maximum number of active users on the map. Thus, it is possible to enable or disable collaboration.
Given text	Provide a text that can be used for internal micro-references between the text and argument parts.

Table 6.2: Typical user interface and context options of existing argumentation systems defined via the template in LASAD

6.2.4 Definition of Argumentation Sessions

The last part of the configuration is the definition of instances that are actually used to argue. These instances are referred to as argumentation sessions. A session is created based on a template, which is in turn based on an ontology. Each session can be restricted in visibility to one or multiple users or roles. Thus, typical research and teaching settings can be applied. By means of multiple sessions, single script activities can be implemented.

The session concept can be exemplified by means of a classroom setting. Here, 30 users are required to solve an argumentative task on their own. Thus, the teacher would define the underlying argument model (ontology) and the user interface components that are available (template) once. After that, the teacher will create 30 sessions (using this ontology and template configuration), that is, one for each student. Thus, each student will use the same underlying argument model model and the same user interface components while — at the same time — working on his or her own argument. In the context of a scripted approach, a user might be required to generate multiple arguments. An example is a two-step approach in which a user is required to first create an argument on his own and, afterwards, create an argument together with another user based on the same configuration. Here, three sessions will be created for each dyad. Two of them would be restricted to the single users and, hence, only visible to the specific user. The third one, however, will not be restricted and, thus, available for both of them. By means of the session concept, the configurations have to be done only once and can then be used multiple times.

6.3 Logging Mechanisms

As mentioned in section 6.1.3, the framework supports multiple ways of logging: (1) state-based logging, (2) action-based logging, and (3) hybrid logging. While existing argumentation systems usually focus on one of the ways, there are reasons to support all of them. This will be discussed in this section.

6.3.1 Action-based Logging

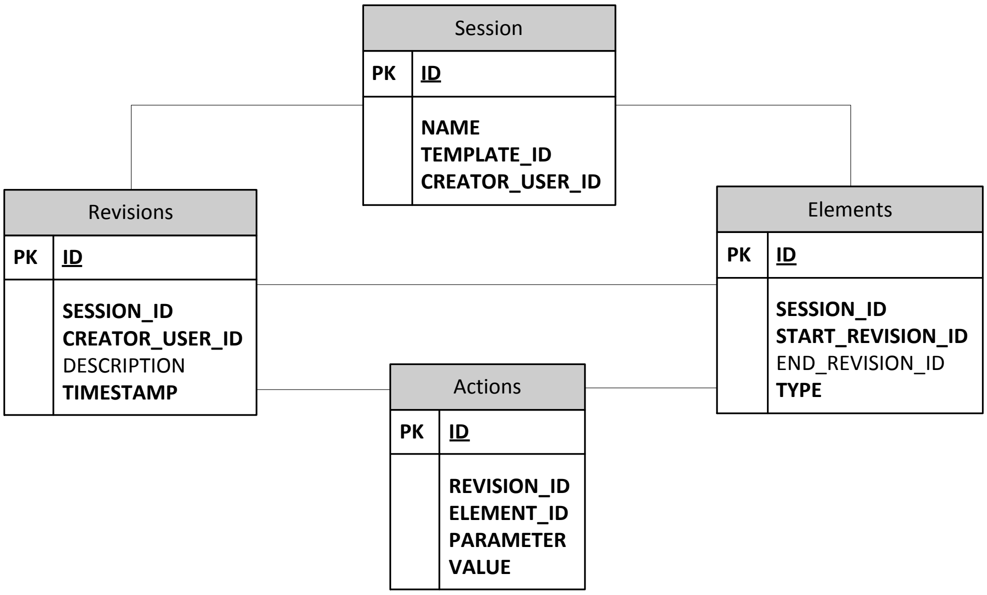


Figure 6.4: Extract of database schema to enable action-based logging

The action-based logging is the standard way to persistently save the results of the argumentation processes in LASAD. Each user action which results in a persistent change of the argumentation — for instance, the adding or deletion of elements to or from an argumentation session — will initiate a new revision of the argumentation session. As shown in Figure 6.4 each revision is connected to an element and a set of actions that actually modifies this element. The actions that are connected to a revision consist a set of parameters, that is name-value pairs (e.g., *TEXT (name) - "In my opinion..." (value)*). Further, each element has a start as well as an end revision (once it is deleted), which describe the element's scope. By means of this concept it is easily possible to replay the whole argumentation process step by step, that is going through the revisions.

6.3.2 State-based Logging

While the action-based approach is beneficial for purposes in which the argumentation process is of interest, for instance in research or teaching, it is problematic from a performance perspective, in particular when considering longer argumentation sessions. In order to join the argumentation later, the complete argumentation process has to be replayed to get the current state. While this works fine in the beginning, the required time and server workload increases the longer the argumentation session lasts.

Here, state-based logs are preferable. That is, only the current state of the argument map will be saved (for instance in XML). When connecting these state descriptions to the underlying configurations, that is, the ontology and template definitions, other instances of the framework are also able to import these files and configurations. In conclusion, the use of state-based logs improve the performance for later joins at the cost of losing the process details.

6.3.3 Hybrid Logging

A third approach that can be created from action-based logs is a hybrid log. Here, process details that are no longer relevant for the current state, for instance, the creation of elements that have been deleted at a later stage would be filtered out before sending the replay details to the client. This is achieved by means of a versioning[1] of the argument.

To exemplify this concept: Imagine there is an argument consisting of three parts A, B and C, which are all added in version 1 of the argument. In version 2, part C will be deleted. In version 3, part D will be added. After that, a new client joins the argumentation. At this point, a pure action-based log would send the complete process to the client. A hybrid approach, however, would skip the creation and deletion of part C, because it is not important to get to the latest version.

By means of such a hybrid approach, the workload of a client that does not require the complete process will be reduced and approximated to a state-based log (even though a state-based log would still be of higher performance caused by additional overhead of the action sequences). At the same time, it would be possible at any point to get the complete process if required. Thus, a hybrid logging approach is beneficial compared to a pure action-based or state-based log.

[1] Versioning is used in many computer science contexts. An example is Apache Subversion (http://subversion.apache.org/).

7 The LASAD Framework - Reference Implementation

To show that the proposed architecture, presented in chapter 6, is capable of dealing with the requirements collected in chapter 5, the next step consisted of the concrete development of a software framework that implements the theoretical concepts. In this chapter, the used technologies as well as the decisions that led to the choice to use this technology will be explained.

7.1 Client

On the client's side, a web-based approach was considered to fit best to the requirements. The key aspect here was the avoidance of any installation procedure: Each modern computer provides a web browser that enables the user to use web-based applications (→ *NFRs 5 & 7: Usability & Motivation; Platform Independence*). However, classical static web pages were not sufficient for the framework's purpose. Thus, the preferred method was the use of a Rich-Internet-Application (RIA) framework. According to Fraternali et al. (2010), a RIA can be understand as following:

> *"The term RIA [Rich Internet Application] refers to a heterogeneous family of solutions, characterized by a common goal of adding new capabilities to the conventional hypertext-based Web. RIAs combine the Web's lightweight distribution architecture with desktop applications' interface interactivity and computation power, and the resulting combination improves all the elements of a Web application (data, business logic, communication and presentation)." (Fraternali et al., 2010, p. 10)*

However, there is a vast number of RIA frameworks based on different technologies available. In addition, these technologies have major implications for practical use which can be used for further differentiation and categorization of the frameworks:

Place of code execution The first categorizing factor is the location of execution. There are basically two concepts: client-based execution (e.g., Adobe Flash[1],

[1]http://www.adobe.com/de/software/flash/

Microsoft Silverlight[2], Java FX[3], Google Web Toolkit[4], OpenLaszlo[5]) and server-based execution (e.g., Apache Wicket[6]). Whereas the latter benefits from extended capabilities of languages that are designed to be run locally (such as Java or C++), it is detrimental for the client's interface fluency: Each action done on the client's side, even low-level ones such as opening menus, clicking buttons, etc., must be sent to the server, processed by the server, and sent back to the client. Thus, the client is largely dependent on the connection speed. Further, the server will be faced with additional load that can be avoided in a client-based execution approach. Here, the server provides the information required on start-up. After that, all actions that can be executed locally (e.g., all user interface connected actions), will not require any server communication. Hence, it leads to a reduced network and server load, because client resources are used.

Installation requirements The second distinguishing factor, which is somewhat connected to the place of execution, is installation requirements. Here, three approaches have been established:

1. *Installation of local runtime environments* such as the Java Runtime Environment (JRE) as used, for instance, for JavaFX applications. In this approach, there are nearly no limitations concerning the language features, but the advantages are bought at cost of a local installation.
2. *Installation of browser plug-ins* such as Adobe Flash or Microsoft Silverlight. In this approach, the limited capabilities of the web browser are extended by a set of additional functions defined by the plug-in languages. In comparison to the first approach, the installation is less extensive, but still required.
3. *No additional installation required* such as Google Web Toolkit or OpenLaszlo. In this approach, the application can only use the browser's capabilities for visualization and interaction. The clear benefit here is that no installation (beyond a web browser which is usually available on modern computers) is required at all.

In Appendix D a short overview of RIAs is given. However, the amount of available RIAs would fill a book by itself. Therefore, only the most important ones will be presented there. These RIAs can be understood as representative for their categorization. Other framework within the same category would stand out or suffer from similar strength and weaknesses.

In the concrete case of LASAD, the avoidance of any installation process in combination with platform independency had top priority (→ *NFRs 5 & 7: Usability & Motivation; Platform Independence*). To avoid high server costs, the majority of workload should be handled by the clients. Thus, the final decision was made between the

[2]http://www.microsoft.com/silverlight/
[3]http://javafx.com/
[4]http://code.google.com/webtoolkit
[5]http://www.openlaszlo.org/
[6]http://wicket.apache.org/

RIA frameworks that do not require any installation and were client-focused. Here, the candidates were OpenLaszlo and Google Web Toolkit (GWT). The final decision was made in favor of GWT. In GWT the basic programming is done in Java, which is then compiled into JavaScript. According to the TIOBE Index[7] (a programming community index based on search results of popular search engines), Java was in November 2008 (when the decision was made) and even today (July 2011) the most popular programming language. In addition, the LASAD project team was familiar with it and the community behind GWT was bigger than the community of OpenLaszlo, which were considered helpful to solve problems in the development phase.

7.2 Server

In order to keep the barrier for future extensions of the reference implementation as low as possible, the server was expected to use the same programming language as the client. Thus, a Java based server implementation was favored. A first prototype of the server was built using Enterprise JavaBeans (EJB), which were expected to be able to deal with a potentially high workload since EJBs are typically used in large business applications. However, in order to run EJBs a web application server such as, for instance, JBoss[8] or GlassFish[9] is required. A direct consequence of the use of a web application server is an increased complexity concerning the installation process (→ *NFRs 5 & 7: Usability & Motivation; Platform Independence*). Together with unsatisfying performance due to additional overhead caused by the EJBs, the server was re-implemented in plain Java. In order to keep the platform independency of the server side, there were two communication interfaces implemented (→ *NFR7: Platform Independence*). On the one hand, a Java Remote Method Invocation (RMI) interface. This allows easy communication with external Java based clients (→ *NFR3: Extensibility*). On the other hand, a web-service interface in order to communicate with clients not using Java (→ *NFR7: Platform Independence*). A proof of concept concerning the platform independence of the reference implementation by means of web-services will be presented in chapter 9.1.3.

However, GWT is neither able to directly communicate with the server via RMI nor via web-services. Instead, GWT provides the possibility to call Servlets, which are defined in the GWT project. Thus, the communication shown in Figure 7.1 has been implemented to enable communication between GWT and the Java server.

The basic concept here is the forwarding of information from GWT via the Servlet to the server. For non-GWT clients, the Servlet layer can be dropped. Here, the server is communicating with other clients either in form of a server push approach via a RMI or web-service call from the server to the potential client. In order to keep flexibility up, the clients are also allowed to poll, that is, to ask the server regularly for updates.

[7] http://www.tiobe.com/index.php/content/paperinfo/tpci/index.html
[8] http://www.jboss.org/
[9] http://glassfish.java.net/

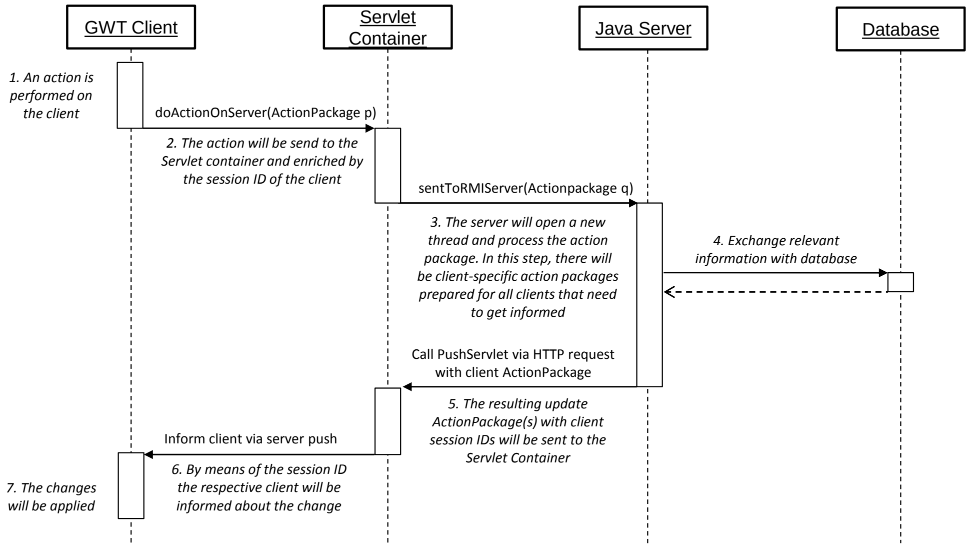


Figure 7.1: Communication sequence in Google Web Toolkit implementation

7.3 Database

On the database layer, it was important to use an open and freely available tool (→ *NFR6: Openness*). The requirements on the data layer are low — just the storage of configuration files as well as logging of information. These are easily fulfilled by relational database management systems which are typically applied in practice. However, there are alternatives that might be considered. Here, the rather new approach of NoSQL (Not only SQL) databases should be mentioned. These include key-value stores (e.g., Amazon's SimpleDB[10] and Dynamo (DeCandia et al., 2007), Scalaris (Schütt et al., 2008), MemcacheDB[11]), column-oriented databases (e.g., Facebook's / Apache's Cassandra (Lakshman & Malik, 2010), Apache's Hbase [12], Google's Bigtable (Chang et al., 2006)), document-based stores (e.g., Apache's CouchDB[13], MongoDB[14]) as well as graph-based databases (e.g., neo4j[15], Sones GraphDB[16]). A detailed list of existing NoSQL databases can be found at http://nosql-database.org/. NoSQL databases are typically used to face the challenges caused by heavy load Web 2.0 applications such as Facebook. Assumed benefits of the NoSQL approach include a faster processing of data as well as a higher degree of flexibility (Leavitt, 2010; Stone-

[10] http://aws.amazon.com/simpledb/
[11] http://memcachedb.org/
[12] http://hbase.apache.org/
[13] http://couchdb.apache.org/
[14] http://www.mongodb.org/
[15] http://neo4j.org/
[16] http://www.sones.com/

braker, 2010; Vicknair et al., 2010). In addition, the inquiry language SQL is replaced by more intuitive concepts.

However, the performance benefits depends on the concrete needs of the project and cannot be generalized. In addition, the LASAD framework is unlikely to be required to deal with workloads similar to Facebook or Amazon. Thus, the decision within the reference implementation was made in favor of a classic relational database, which is also open-source and, thus, freely available — mySQL[17]. The clear benefit is the avoidance of unforeseen errors which are likely to occur in early releases of new approaches, which NoSQL databases clearly are. Further, most computer scientists are familiar with SQL so that this is not a barrier at all.

7.4 Authoring Tool

To enable users that are unfamiliar with programming to benefit from the framework's flexibility, the reference implementation includes a visual editor to configure the system. This authoring tool follows the multi-step approach presented in Figure 7.2, which is designed to guide the configuration process.

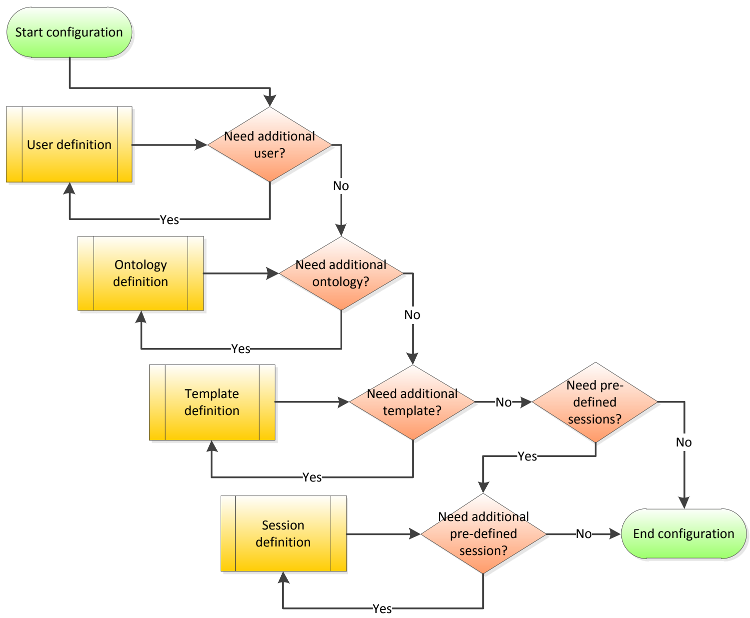


Figure 7.2: Flowchart of the multi-step approach of the authoring process

[17]http://www.mysql.com/

The authoring process is strictly oriented at the basic configuration concepts of the framework described in chapter 6.2. That is, the steps include a definition of user accounts, a definition of the ontology, that is the underlying argument model, a template definition which specifies the available user interface element as well as the argumentation context and the definition of sessions. Each of these steps can be configured independently, that is if one only wants to add a new template based on an existing ontology, the ontology definition can be skipped. Overall, the authoring process results in XML configuration files. A screenshot of the authoring tool is given in Figure 7.3.

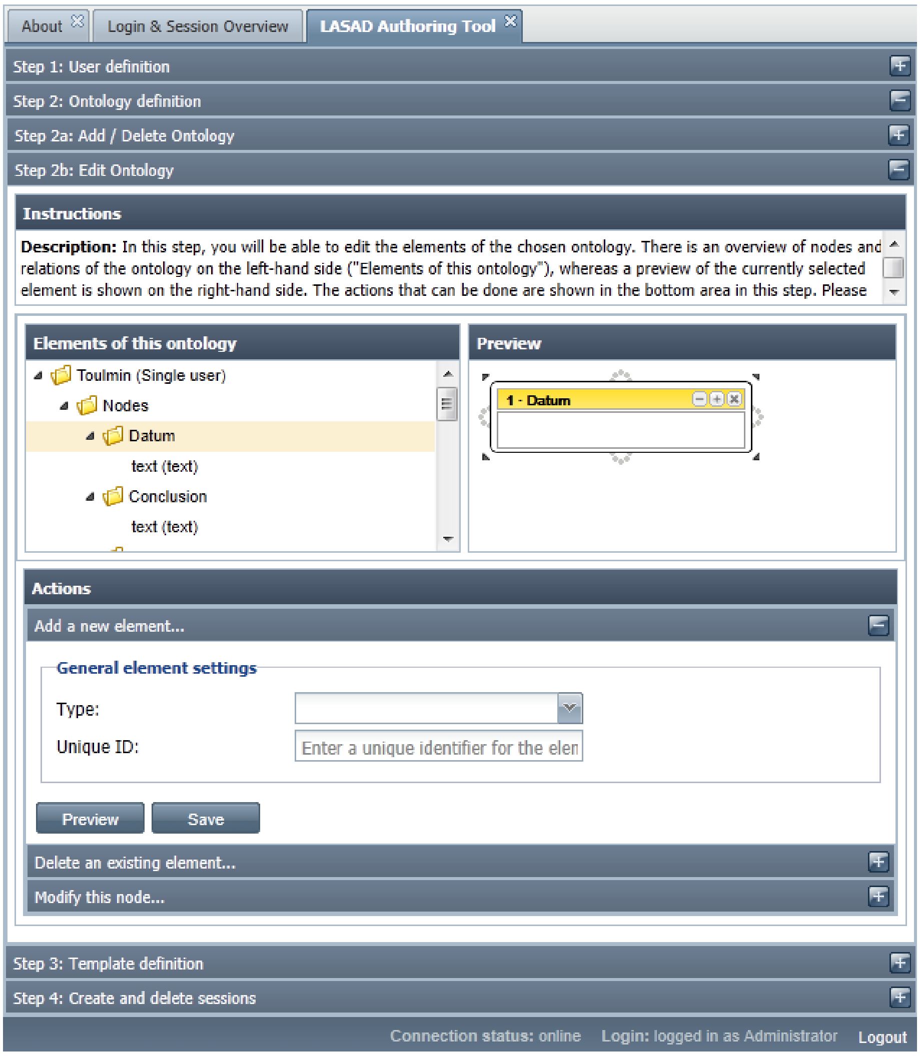


Figure 7.3: Screenshot of the LASAD authoring tool

However, the authoring tool currently only supports the basic configuration mechanisms of the framework. Not supported is, for instance, the configuration of external analysis and feedback clients (→ chapter 9.1.3) or the use of multiple visualizations (→ chapter 9.1.2).

7.5 Challenges and Solutions

In chapter 4, it was mentioned that the motivation behind design decisions as well as potential failures are of high value for future development. Thus, I would like to present some examples of challenges that had to be solved during the development process of the reference implementation.

7.5.1 Drawing Operations in Web Browsers

Problem Even though GWT is well suited to deal with everyday requirements for most web applications, it is still limited in some ways. One of these limitation is the missing support for drawing operations[18]. Thus, drawings had to get implemented without the help of the GWT framework.

Solution In order to support drawing operations, the open source library gwt-diagrams[19] has been used and extended to deal with LASAD specific requirements. The basic concept is the use of different drawing operations for different web browsers. Based on the browser, the adequate implementation will be chosen during runtime. These implementations include the drawing via Canvas (for most browsers such as Firefox, Safari, Opera) and VML (for Internet Explorer).

Alternatives In newer versions of the GWT framework, drawing by means of HTML5 constructs is possible. Even though not supported by older web browsers, future developments would clearly use this implementation.

7.5.2 Server Push

Problem Description In order to work collaboratively on a problem it is obvious that all additions and modifications of the argument must be distributed to all group members. However, typical web pages follow a request-respond pattern, that is, in order to retrieve data from the server, one has to explicitly ask for it. In a collaborative environment this pattern will cause problems. First, requesting information manually is annoying. Even though this can be fixed by regular automatic requests, the second problem is still present: How does a user or the system know when new information such as new nodes in a graph-based argument

[18] At least GWT was not supporting them when the development started.
[19] http://code.google.com/p/gwt-diagrams/, written by Michal Balinski

visualization are available? To avoid the ignorance of important data provided by other users, the requests must be sent regularly in a short period of time, which would result in a lot of (in most cases) unnecessary traffic between client and server. As a direct consequence, the server must be more powerful than actually required just because it has to handle a high amount of unnecessary requests.

Solution To solve this problem, a server push functionality is required. This means that the server must be able to notify clients about changes by either sending the changes directly or by sending a note that new data is available. Unfortunately classic web technologies do not provide support for such a functionality. In LASAD this problem was solved by using a Comet (Bozdag et al., 2007; Russell, 2006) implementation. Comet is a technique which emulates a server push mechanism by means of a long polling approach as shown in Figure 7.4.

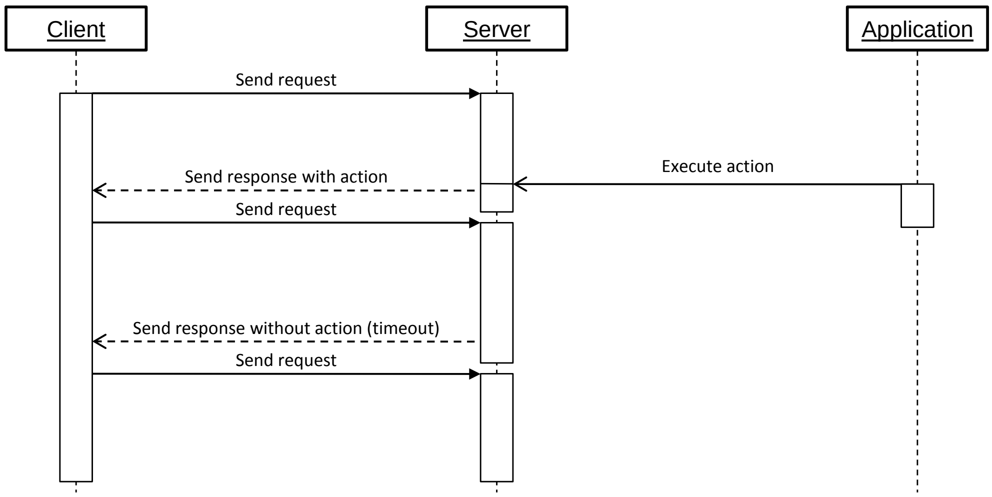


Figure 7.4: Comet server push via long polling sequence illustration

Long polling implies that there are regular requests to the server. In contrast to classic polling, however, the main difference is that the connection will stay open for a longer period of time. In detail, the server will not send a response to the client's request until either data is available (first scenario in Figure 7.4) or a defined maximum time frame is exceeded (second scenario in Figure 7.4). The benefit is obvious: Instead of answering a vast amount of requests, the overall number of requests will be reduced at the cost of open connections. The latter, however, is less performance intense than the former. Thus, the overall approach is beneficial. In LASAD, the external open source library GWTEventService[20] is used as Comet implementation.

[20]The GWTEventService was written by Sven Strohschein. Further information about the library can be found online: http://code.google.com/p/gwteventservice/

Alternatives & Limitations The Comet approach can be only applied to techniques which allow asynchronous client server communication. Otherwise the client's request would block all other actions. In addition, some web browsers limit the maximum number of open connections which could cause problems when using Comet in addition to multiple other requests. During the development of LASAD, the specification of Java Servlet 3.0[21] has been finalized. A native server push functionality is part of the specification and could be considered for future developments.

7.5.3 Balancing Traffic and Workload

Problem The communication sequence shown in Figure 7.1 in section 7.2 is highly dependent on the network connection. While this approach works fine for basic actions such as the creation or modification of argument parts, it is problematic for traffic or workload heavy operations. One of the bigger problems has been caused, for instance, by the mechanism to track the cursor movements of collaboratively arguing users. Each cursor movement was resulting in an action, which was sent to the server. Next, the server creates a revision in the database in order to allow for a replay afterwards, before forwarding the updates to the clients which are active in the argument session. This caused extensive workload for all components of the framework: (1) The client had to send and retrieve a lot of action packages only to update cursor positions. (2) The server had to forward high amounts of packages to the respective clients. (3) The database suffered from extensive load to allow the tracking of cursor movements.

Solution The cursor movements are essential for two purposes. The first one is the awareness. By means of other arguers' cursors, each arguer is aware who is working at which part of the argument session. The second one is research. Here, it might be of interest for certain experiments to identify afterwards which arguer has been working on which part of the argument part when chatting with others since gestural deixis is frequently used. An example is the following chat episode:

> *Ted: Do we have a node that mentions the influence of cars on global warming?*
> *Barney: Yes, it is over here!*

In this example Barney is using his cursor to point to a specific part of the modeled argumentation. In this case, it is essential for Ted to see Barney's cursor (awareness). In addition, when evaluating the argumentation process afterwards, it might be interesting to know about which part of the argument graph they were talking at this point. Thus, the cursor position in the replay is essential as well (research). However, while not a huge problem in typical

[21] The detailed description of the specification can be found online: http://www.jcp.org/en/jsr/detail?id=315

desktop applications, in which the resulting data will be stored locally, it is for networked environments. This fact is amplified by the the generality of the framework, which results in a lot of overhead caused by the generic concept of actions and action packages.

To solve this problem, the following compromise was implemented. First, the client's rate to submit their cursor position was modified. Instead of sending all cursor movements, they will be only sent ones every 1.5 seconds under the condition that the cursor movement since the last submitted update changed by more than 25 pixel in one of the directions. Second, the server was extended by a mechanism that passes actions just along to the other clients. This way actions that are likely to be irrelevant can be dropped from the database logging mechanism. However, in order to solve the problem caused by the research interest, the third adjustment was the persistent logging in the database. That is, the cursor position will only be stored to the database if the update of the cursor movement is submitted within a range of 10 seconds after the last chat message of the respective client.

This trade off had shown to be effective. The update rate was sufficient, the network traffic was reduced and the workload of the database was minimized while, at the same time, fulfilling its purpose.

Alternatives A potential alternative is the use of a file logging mechanism instead of a database oriented one. By means of timestamps it would be possible to reconstruct the respective cursor position at the required time. While this could be done on the server side instead of the database logging mechanism, it would not solve the problem of network traffic. Thus, a client side logging would be more appropriate. Nevertheless, this would cause additional time to collect the data from the machines if needed and it would not be feasible possible to do so for spatially separated computers. To solve this problem, the client side logs could be submitted regularly to the server via the usual action and action package communication structure. However, this solution would not solve the problem that other clients need live updates of cursor positions.

7.5.4 Concurrency Management

Problem In collaborative environments where people will work together at the same time on a shared artifact, modification conflicts are hardly avoidable. Thus, they must be handled in an adequate way.

Solution In order to avoid potential conflicts locks have been used. However, approaches such as locking the complete argument map would not be beneficial for the overall collaboration process (unless one actually intends to use such concept as done, for instance, in the study of Schwarz & Glassner (2007), who used a turn-taking concept). Thus, the extend of locks should, on the one hand, be as fine-granular as possible, but, on the other hand, a high traffic between clients just to manage potential locks should be avoided. Thus, the decision was in favor of an element level locking. However, the only element which required a concrete lock is the text field, because the editing time lasts for a longer period of time. On the server side, these locks are handled using a queue, that is a first-in-first-out concept was applied. In practice, this means that all incoming actions will be processed in the order they reach the server. If one of the actions is an locking request, the server will inform all other clients which participate at the argumentation session that the element is locked now. However, if a concurrent lock or change request arrives, the server will deny the request until the lock is freed by the originator or the originator leaves the argumentation session. This is illustrated in Figure 7.5.

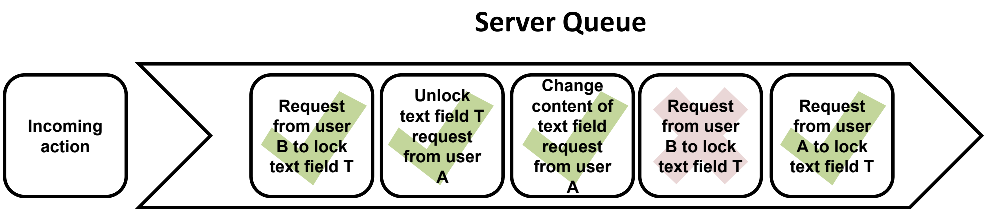


Figure 7.5: Illustration of server queue

Alternatives There are basically two imaginable improvements. The first one is the use of an increased granularity of locks. Instead of blocking, for instance, a complete text element, it might be possible to use word-level locks. The second one is the use of optimistic methods, which would avoid any locks. However, in order to evaluate the appropriateness of these methods, the required traffic and workload must be kept in mind.

7.6 Summary & Outlook

Based on the architecture presented in chapter 6, a web-based reference implementation has been developed. Together, chapters 6 and 7 showed how the requirements of a generic argumentation system can be satisfied. The results are summed up in Table 7.1 and enriched with forward references to examples in some cases.

ID	Criterion	Fulfilled by
1	The functionalities of the system must be modular.	All available elements are modular. This includes the user interface elements (configured via the template, → chapter 6.2.3) as well as the elements of the underlying argument model (configured via the ontology, → chapter 6.2.2).
2	The configuration of the functionality combinations should be done in a central place	All configurations will be stored in a central database (→ chapter 6.1.3).

ID	Criterion	Fulfilled by
3	The modular set of functionalities should be easily extendable on the code level. This should be achieved by means of a common interface.	The available set of functionalities, that is, the elements of the underlying argument model and the user interface elements, is based on a generic parameter concept. These parameters can be manipulated via well-defined actions. The parameters and actions are implemented as strings, which ensures generality and potential for future extensions (→ chapters 6.1.1 & 6.1.3).
4	The framework should be able to communicate with external applications.	On the server layer, there is a generic interface that can be used to communicate with the system via platform dependent concepts (Java RMI) as well as via platform independent concepts (Web services) (→ chapters 6.1, 6.1.2 and 7.2). Examples will be presented in chapters 9.1.3 (for an external client using Java RMI) and 9.1.3 (for an alternative client using Web services).
5	There should be no noteworthy lag between frequently executed users' actions and the system's response (that is, usually less than 1 second).	To avoid unnecessary lags, server push techniques have been implemented (→ chapter 7.5.2). The server was re-implemented in plain Java to improve the overall performance (→ chapter 7.2). In addition, the framework is using a lightweight action format (→ chapter 6.1.1).
6	The framework must scale up in an adequate way (with respect to performance) to support typical classroom sizes of uses (approx. 30).	In addition to the mechanisms described in criterion 5, the general layer architecture (→ chapter 6.1) of the system allows a load distribution that is beneficial for the overall performance. This way, the scalability is ensured.
7	The framework should make use of modular concepts in order to allow a spatial distribution of system components. These components should communicate via a well-defined interface.	The framework is based on a layer architecture (→ chapter 6.1). Each layer has a well-defined interface.
8	The user interface of the system should be intuitive, that is, the amount of required training should be minimal (not more than 15 minutes).	The user interface design was done based on a careful review of existing approaches that have shown to be intuitive (→ chapter 3). The studies that will be presented in chapters 8 and 9 will confirm that the reference implementation is easy to use. In order to enable even inexperienced users to configure LASAD to their needs, the configuration of the framework is supported by an authoring tool (→ chapter 7.4), which will be evaluated in chapter 8.2.
9	No installation should be required on the client side.	The reference implementation of the client is using GWT (→ chapter 7.1). This way, the client runs in a web browser, which is an integral part of each modern operating system. Thus, no client-side installation is required.
10	The installation requirements on the server side should be minimal.	The server is based on plain Java (→ chapter 7.2), which only requires the Java Runtime Environment. However, the non-installation requirement on the client side (see above) had a higher priority. Thus, a tradeoff was made on the server side — here an additional webserver, for instance Apache Tomcat, is required to provide the clients with the required data.

ID	Criterion	Fulfilled by
11	External libraries and technologies used in the framework as well as the framework itself should be open-source or at least freely available.	All technology used is either open-source or freely available (→ chapter 7). In addition, the complete LASAD framework will be made available open-source as well.
12	Internationalization should be supported by the framework.	Internationalization is an integral part of GWT, which is used on the client side (→ chapter 7.1). The reference implementation currently has support for English and German. Future extensions can be easily done.
13	The framework should work platform independent.	The client runs in a browser (→ chapter 7.1). The server requires a Java Runtime Environment (→ chapter 7.2). Both is platform independent.
14	The framework should support multiple visualizations.	The architecture is designed to support multiple clients. Each client can support one or more visualizations of the data (→ chapters 6.1 & 6.1.1). This will be shown in chapters 9.1.2 and 9.1.3.
15	The data format should be flexible to enable visualization independent data exchange as well as future extensions.	The used data format is independent of a concrete visualization. The general concepts of actions and parameters is highly flexible to simplify future extensions (→ chapter 6.1.1).
16	The system should provide a common interface for the data exchange.	The single layers of the architecture communicate via a well-defined interface. Via these interfaces, highly generic action objects are exchanged (→ chapters 6.1 & 6.1.1).
17	The underlying argument model should be modularly configurable.	The underlying argument model is configurable by means of the ontology definition (→ chapters 6.1.3 & 6.2.2).
18	The framework should provide a set of elements that are typically used in argumentation systems (text fields, internal and external references, numeric assessments, choice dialogues).	There is a set of ontology elements available (→ chapter 6.2.2). This set can be extended if required (as done for instance in the Metafora project that will be presented in chapter 9.2.5).
19	The graphical user interface should be modularly configurable.	The user interface is configurable by means of the template definition (→ chapters 6.1.3 & 6.2.3).
20	The framework should provide a set of user interface parts that are typically used in argumentation systems.	There is a set of user interface elements available (→ chapter 6.2.3). This set can be extended if required (as done for instance in a study that will be presented chapter 9.2.3).
21	The framework should have a scripting engine that allows the definition and execution of collaboration and learning scripts.	Currently, the framework does not have a complete scripting engine integrated. However, the basic parts of typical scripts (for instance, different visualizations and different collaboration settings involving different roles) are supported. An example of a manual scripting approach, which makes use of this support will be presented in chapter 9.2.3. In addition, the architecture is prepared to integrate a scripting engine (→ chapter 6.1).

ID	Criterion	Fulfilled by
22	The framework should have a roles and rights management.	A simple version of a roles and rights management is implemented as part of the user definition (→ chapters 6.1.2, 6.2 & 6.2.1).
23	The provision of feedback from multiple sources (internal and external) should be supported.	The framework's architecture supports multiple clients (including feedback clients) (→ chapter 6.1). Each feedback agent can act in the same way as regular user. That is, any feedback agent will get a specific role with certain rights assigned. Further, human moderators are able to participate at the argumentation process with additional rights via the same client as the other users.
24	It should be possible to watch a replay of the argumentation process based on action-based logs.	The action-based logging mechanism (→ chapter 6.3.1), which is used in the framework, is sophisticated enough to allow a complete replay. Therefore, a replay function has been integrated into the reference implementation.
25	The framework should support the creation of snapshots in form of state-based logs.	Based on a hybrid logging approach (→ chapter 6.3.3), state-based snapshots (→ chapter 6.3.2) can be done.
26	There should be an export/import function to exchange arguments.	The state-based snapshots can be exported into XML and exchanged between clients (→ chapter 6.3.2).
27	The framework should provide communication facilities to support arguing in groups.	There is a chat function, which can be configured as part of the template definition (→ chapter 6.2.3).
28	The framework should provide awareness aids to the users to promote a fluent collaboration (meta-data indicating who created which part of the argument and when as well as who is currently working on which part).	Awareness information can be provided on multiple levels. On the one hand, there is the ontology level. By means of a configurable ontology element the author of elements can be shown to all users (→ chapter 6.2.2). On the other hand, there is the template definition. Here, additional tools to support group interactions can be configured such as showing cursor positions of other users or providing a group pointer (→ chapter 6.2.3).
29	In collaborative settings, there has to be a concurrency control that ensures that a data loss as consequence of concurrent access is avoided.	The concurrency control is done on the server layer of the framework (→ chapter 6.1.2). In the reference implementation, the concurrency control uses an action queue. The use of locks is minimized so that they are only applied to time consuming actions such as editing the text of an argument part (→ chapter 7.5.4). This way, a fluent collaboration is ensured.

Table 7.1: Overview of fulfillment of functional and non-functional requirements in LASAD

The overview show that the requirements collected in chapter 5 have been fulfilled. In addition, it confirms the implicit assumption present in research question 3 (*How can the divergent requirements and expectations of various domains and multiple groups be met by a single framework?*) that it is possible to create a framework that is able to satisfy domain-specific needs at least from a theoretical point of view. Concerning the practical side, the following chapters will present actual use of the framework.

However, a software development process is a never-ending story. This is also true for the reference implementation of the presented framework. That is, some of the requirements are fulfilled on a conceptual level, but only implemented rudimentarily and will evolve in the next years. An example is the support for collaboration and/or learning scripts. The proposed way to realize these scripts, is the combination of multiple system configurations. However, in the current version of the reference implementation the process management, that is, the decision when which configuration is used, has to be handled manually. In future versions of the reference implementation, an engine which automates these steps will be integrated.

While functionality extensions are valuable, the usability aspect must be kept in mind at all time. Here, a first step in form of an authoring tool for the basic configuration of the framework has been taken. Nevertheless it is important to keep the entry threshold as low as possible and enable inexperienced user groups to benefit from the functionality LASAD offers. Thus, future extensions of the authoring tool should involve the definition of concrete script definitions that can be executed in the tool and the option to integrate external clients in the argumentation process. The former is a direct consequence of the functionality extensions. Here, an imaginable way to realize the authoring of scripts is to use a specific configuration of the system. Therefore, the ontology could compromise typical elements that are used in scripts such as phases, activities, roles, etc. (see chapter 3.5 for details). An illustration of such a configuration is shown in Figure 7.6. The latter, however, is needed to include alternative visualizations as well as analysis and feedback mechanisms that can be applied in certain steps of the argumentation. These steps can be part of scripts and, hence, is directly connected to the scripting capabilities.

Overall, chapters 6 and 7 provided answers to the third research question of this thesis and highlighted the suitability of the generic approach that has been used in this context.

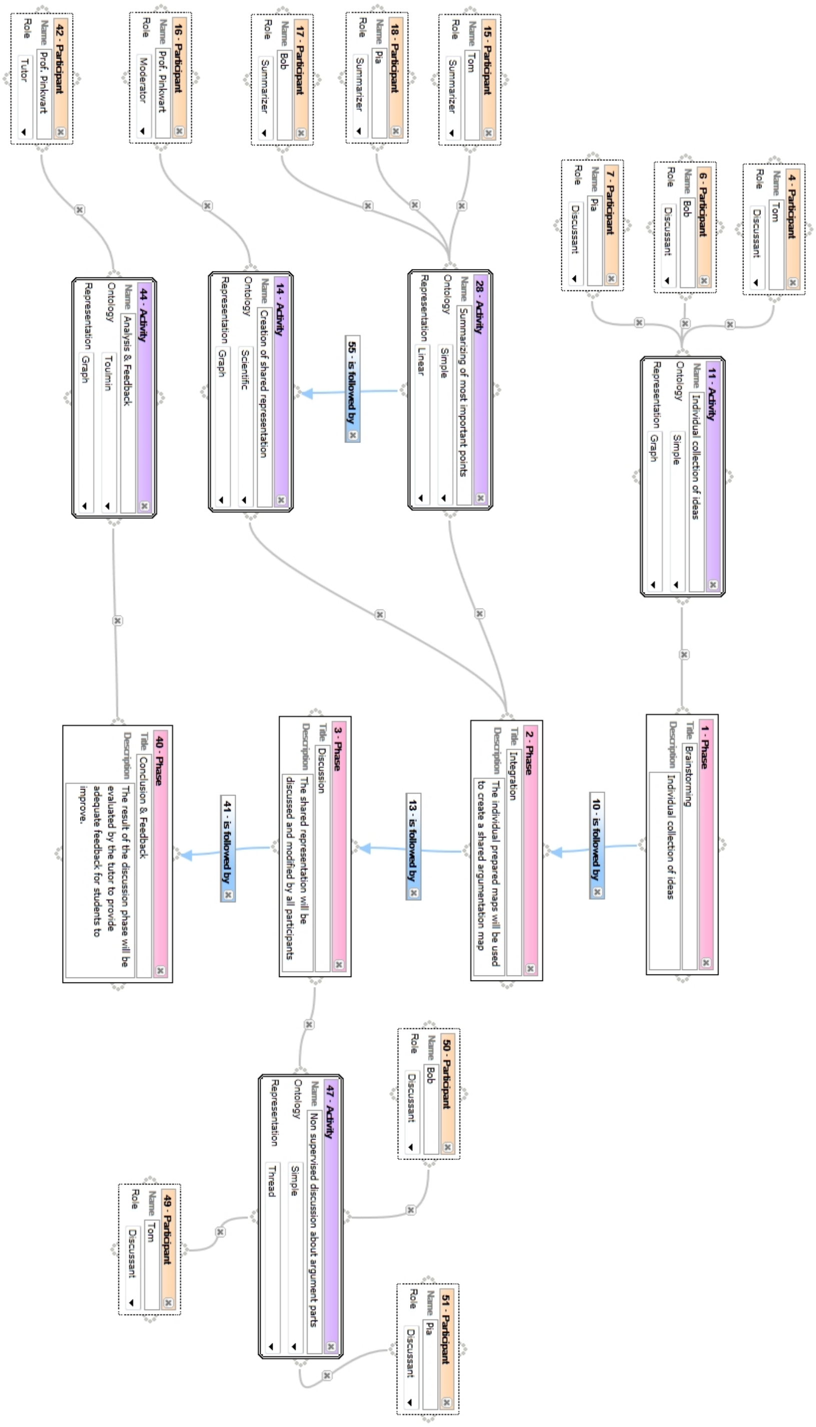


Figure 7.6: Script definition example

8 Internal Evaluation of the LASAD Framework

To prove the system's capability to fit domain-specific needs, a two-tier approach will be used including internal and external evaluations of the framework. In this chapter, the internal evaluations will be described. These describe studies using the framework conducted by myself. The main purpose of the studies is to highlight the applicability of the framework to concrete research challenges on the one hand (cf. research question 4 in chapter 1.2), but on the other hand, show that even inexperienced users are able to use the framework (cf. research question 5 in chapter 1.2). The second tier of the framework's evaluation will be presented in the next chapter. Here, external applications of the framework will be described. This way, it will be shown that the framework is grown-up, that is, other users are able to adequately use it for their purposes and, hence, that it is capable of serving as research and teaching tool.

In this chapter, two studies will be described. The first study that has been published in (Loll & Pinkwart, 2011), presented in section 8.1, will highlight the suitability of LASAD to be used as research tool. In this study, the role of different ontologies and collaboration on the outcomes of argumentation will be evaluated. In particular, I will look for possible interaction effects between them. As the review of existing argumentation systems revealed (see chapter 3), this is still an open issue. By means of this study, I will provide a proof-of-concept, that is, I will show that the framework is flexible enough to be used in a typical research scenario.

The focus of the second study, presented in section 8.2, is the test of the authoring part of the framework. Here, I will show that users who are unfamiliar with programming or even argumentation tools are able to configure the system to their needs. This is important to make sure that, for instance, teachers and students are able to use the framework successfully for their purposes.

8.1 The Impact of Ontology and Collaboration on the Outcomes of Computer-Supported Argumentation

The empirical results presented in chapter 3.9.3 rose doubts about the beneficial role of collaboration. Nevertheless, most experts in the survey agreed that collaboration is essential in many aspects of argumentation, even though they agreed that there are typical mistakes that are likely to occur in collaborative settings.

A similar point is the use of ontologies. Even though there is evidence that the presence of an ontology is beneficial for the outcomes of argumentation (see chapter 3.9.4), an open question is still which ontology is suitable for what kind of task. As shown in chapter 3.6, the number of approaches is huge and there are voices that report problems when dealing with sophisticated ontologies. However, the survey also revealed that most experts use argumentation models, e.g., the Toulmin argumentation scheme, to actually judge the quality of an argument.

The role of these concepts is, thus, not completely clear. In addition, potential interaction effects have not been evaluated as well. When considering, for instance, the controversially discussed topic of over-scripting (Dillenbourg, 2002; Stegmann et al., 2011), one could argue that the use of a highly complex ontology could be detrimental for collaborative settings, because the amount of required coordination in group work is already a highly complex endeavor. On the other hand, a highly complex ontology could be beneficial in assisting group work due to the provided structural means of the ontology.

In order to answer these questions, this chapter will evaluate the impact of ontology and collaboration on the outcomes of computer-supported argumentation.

8.1.1 Hypotheses

The hypotheses that will be evaluated in this study were divided into three categories: effects of collaboration, effects of argument ontologies and interaction effects.

Effects of collaboration

(C1) Arguing in groups (as opposed to constructing arguments individually) will lead to a more elaborated argument, that is, an argument of higher quality, due to different points-of-views of the participants.

In this context, the quality of arguments is seen as a direct result of the successful application of argumentation skills. That is, an argument is of high quality if single points-of-views are well grounded, alternative positions are considered and evaluated by means of adequate facts (cf. chapters 2.1 & 2.2.1). Thus, a more elaborated argument would typically include a higher number of reasoned contributions.

(C2) In collaborative argumentation activities, students will be more motivated than in individual argumentation sessions. This is hypothesized based on the fact that discussions with other arguers will lead to a greater variation of the task steps and, hence, to a less monotonous activity. Prior results by Pinkwart et al.

(2008) highlighted the importance of motivation to promote good learning results in argumentation activities.

(C3) In collaborative sessions, the participation of single users may drop as compared to individual argumentation sessions: shy arguers may stop arguing against a dominant, leading group member and the potential to create free riders (Kerr & Bruun, 1983) increases.

(C4) Collaborative argumentation will lead to more off-topic activities. Prior results by Schwarz & Glassner (2007) show that groups tend to get distracted from tasks, which can be detrimental for the overall argumentation process.

(C5) Group members will review and respond to each other's arguments, and, hence, the overall number of mistakes will decrease in comparison to individual activities. This is hypothesized because argumentation is not a trivial undertaking: users may oversee their mistakes and, by discussing about parts of the argument, typical mistakes of single users may be revealed and corrected

Effects of argument ontologies

(O1) The higher the structural degree of an argument ontology is, the higher the overall structure of the argument map is to be expected. This is a direct consequence if the ontology is used correctly and supported by the findings of Schwarz & Glassner (2007).

(O2) The more detailed an argument ontology is, the more elaborated the resulting argument will be. The rationale for this hypothesis is that the multiple elements of detailed ontologies are expected to prompt the users to make use of them and, hence, think about how to fill them with appropriate materials. Clark & Brennan (1991) also noted that it is easier to refer to knowledge units which have a visual manifestation, so that the presence of various, different ontology elements may lead to more discussions and, consequently, to a more detailed resulting argument.

When considering the existing argumentation systems, this would imply that the very general argument model provided by Athena ("nodes" connected via "pro" and "con" relations) would be inferior compared to a Toulmin based argument model, which explicitly contains structures for more elaborated concepts such as "datum", "warrant" and "rebuttal" (cf. chapter 3.6).

Interaction effects

(I1) For group argumentation, it is hypothesized that the used ontology will influence the degree of collaboration: a more complex ontology may increase the need for collaboration (in order to discuss how to use the different elements to build an argument).

(I2) In collaborative sessions, highly structured argument ontologies may be detrimental to the quality of the resulting argument (due to the double complexity of keeping track of the group process and using a complicated argument model at the same time), while the scaffolds that more structured ontologies provide may be more helpful in individual usage.

8.1.2 Design

To investigate the hypotheses, a mixed 3x2 design was used. The between-subject factor was the argument ontology. Here, the following three different ontologies were used:

1. A simple domain-independent ontology consisting of a general contribution type (*contribution*) and three different relation types (*pro, contra, undefined*). The domain-independency is, hence, achieved by means of a rather unstructured approach.
2. A second domain-independent ontology based on the Toulmin argumentation scheme (Toulmin, 2003). It consists of five different contribution types (*datum, conclusion, warrant, backing, rebuttal*) and four different relations (*qualifier, on account of, unless, since*). In comparison to the first ontology, the universality is achieved by means of a highly structured approach.
3. A domain-specific ontology which was inspired by the Belvedere ontology (Suthers, 2003) consisting of three contribution types (*hypothesis, fact, undefined*) and three relation types (*pro, contra, undefined*). This approach has shown to be effective for scientific argumentation.

The within-subject factor in the study was collaboration. Each participant was required to argue about one topic on his or her own and about another topic in a group of three. To eliminate possible confounds, counterbalancing was used so that half of the participants began with the group phase while the other half began with the single user phase. In the group phase, each participant worked on one machine. The participants were only allowed to communicate via the chat tool integrated in the argument framework. This simulated a remote discussion even though the users were located in the same room (the experimenter was in this room to enforce the rule; in addition the participants did not have eye-contact). Overall, the study took 6 hours per user, including a 1 hour break between two sessions.

8.1.3 Tasks

Each participant worked on two open scientific problems that have no obvious solution. This kind of task choice was motivated by Toth et al. (2001), who used challenging science problems to simulate an authentic argument activity, avoiding a demotivation of students caused by hiding the answer of already solved questions. The Schwarz et al. (2000) results support this decision: their findings include that argumentation is most effective if students are arguing under uncertainty. In this study, the concrete topics for the arguments were:

(Topic 1) The potential of alternative concepts for automotives (including the electronic car, the fuel cell, and biofuel)

(Topic 2) The German energy mixture in 2030 (including nuclear power, fossil fuel, and renewable resources)

For each topic, three different possible positions were prepared. To allow all participants to argue for or against each of these positions, the students were provided with two pages of background information per position. This material, given in form of material chunks (graphs, tables as well as plain text) was typical for scientific argumentation, including facts, examples, statistical data and observations. The concrete material can be found in Appendix E. In addition, there was one page containing material that was common to all positions. The participants were explicitly allowed to go beyond the given material in their arguments. Each session about a topic was split into four slots of 30 minutes each. In each of the first three slots, the participants were given the background material for one of the three positions (e.g., "nuclear power as a future energy") as well as the common materials and were asked to create an argument about this position using the LASAD system. The fourth slot was used to integrate the three separate positions and to draw a final conclusion to solve the argumentation task. For this last step, the participants were given the materials for all positions again.

8.1.4 Participants & Training

Overall, 36 (under-)graduate students (25 male, 11 female) with different majors participated in the study. They were between 19-35 years old ($m = 24.64$, $sd = 3.68$) and in semesters 1 to 22 ($m = 7.00, sd = 5.62$). All participants were either native German speakers or fluent in this language (the complete study was conducted in German). Participation was voluntary and all participants were paid for completing the study. The participants were assigned randomly to all three "ontology" conditions, that is, in each condition there were four groups consisting of three students each. In all but one group was one female student.

None of the participants had used the argumentation system before. Thus, a short video introduction (15 minutes) to the LASAD system was shown to make sure that all participants had the same basis. All videos consisted of three parts: (1) A general introduction how to interact with the system, (2) an overview of supporting features to work in groups (e.g., chat, cursor tracking), and (3) an ontology dependent part in which the condition dependent features of the system were explained using an example common to all conditions. Finally, the example argument that was presented in the video was distributed among all participants on paper and was available during the complete study.

8.1.5 Tests & Interviews

To test the learning effects caused by the argumentation tools use, three multiple-choice tests on argumentation abilities as well as two multiple-choice knowledge tests per topic were used. The tasks of the argumentation tests were taken from a list of questions of the *Law School Admission Test (LSAT)*[1]. Each argumentation ability test consisted of four questions, two from the area of logical reasoning and two from the area of analytical reasoning. These questions were not specific to law, but to argumentation in general. These tests took place before the first session, between the two sessions and after the second session. The order of the tests was counterbalanced. The participants were given 6 minutes (1.5 minutes per question) per argumentation test.

The knowledge tests were centered on the domain of argumentation in the respective study sessions (automotive concepts and energy mix). They were administered immediately before and after the corresponding sessions (in a counterbalanced manner) to measure domain learning. The participants were given 4 minutes (1 minute per multiple-choice question) per knowledge test.

In addition to these two tests, a questionnaire was used to evaluate the usability of the overall LASAD argumentation system. By means of this test, it should be check whether certain features of the system might have hindered the students to engage in reasonable argumentation, especially since this was the first larger study with the LASAD system. Here, the standardized System Usability Scale (Brooke, 1996) which has shown to be an accepted measure for usability (Bangor et al., 2008, 2009) was used.

Finally, the participants were asked in an open interview with the experimenter about their motivation during the study sessions, and about potential problems and ideas for future improvements of the system.

8.1.6 Coding Procedure

The material distributed to the participants (see Appendix E) consisted of unconnected information chunks including relevant as well as non-relevant parts. To be able to check how much of the relevant material was used, three domain experts independently created a list of all the facts that could either be directly taken from the material or directly concluded based on a combination of multiple information chunks. These lists were merged and discussed; the resulting lists (containing 81 entries for topic 1 and 75 for topic 2) were used as a reference for the relevant information that can be extracted from the hand-out material.

To get further insights into the resulting argument maps, 6 of 48 maps (one individual map and one collaborative map for each ontology, that is, 12.5% of all the maps)

[1] http://www.lsac.org/JD/LSAT/about-the-LSAT.asp, last visited: 2011-06-17

were coded element-wise independently by two coders with respect to the use of given material. For each element (boxes and relations) in a diagram, the coders checked if the contained information was based on a fact in the "reference list" (cf. above) or if it was a completely new contribution. The coders also rated the correctness of the used ontology elements (if, for instance, a fact element was actually used to represent a fact). To judge the structural quality of an argument map, the coders additionally checked for each of the 6 chosen maps if this map contains (a) a starting hypothesis, (b) a conclusion, and (c) a clear grouping of the different positions.

Based on these coding results, the inter-rater reliability was calculated and resulted in a Cohen's κ of 0.60 for the material used and 0.61 for the used elements. Concerning the general structural features (a-c), both coders agreed 100% on each measure. Taking into account the ill-defined nature of argumentation (Lynch et al., 2010), this level of agreement can be assumed to be acceptable overall. The remaining elements were then coded by one coder in the same manner as described above. Overall, 5477 elements were manually coded this way.

To measure the degree of coordination, also the chat messages were encoded. First, the chats (consisting of 878 messages) were divided independently by two coders into episodes that belong together, e.g. a discussion about where to start with argument modeling. Slight differences were resolved by discussion between the coders. This resulted in an overall number of 196 chat episodes. Based on the chat episodes of three sessions (one per ontology, that is, 25% of all material), the following four categories were agreed on as a coding scheme for the chat episodes: (1) Content, (2) Structure, (3) Coordination, (4) Off-topic. Based on this coding scheme, each chat episode within the 12 collaborative sessions was independently coded by two raters. The raters achieved a moderate Cohen's κ of 0.56. However, it turned out that the categories "structure" and "coordination" were often not clearly distinguishable so that these two categories were merged into one, which resulted in a high κ of 0.76. The raters resolved remaining conflicts through discussion.

8.1.7 Results

System Usability

This study was the first one done with LASAD framework. Therefore, the general usability of the system was in the focus of interest to make sure that there were no detrimental influences of the system on the other outcomes of the study. The SUS test resulted in a mean score of 81.46 (which is similar to a B grade). The concrete questions of the test that have to be answered on a 5pt Likert scale (1 = strongly disagree to 5 = strongly agree) were the following:

(Q1) Ich denke ich würde dieses Programm gerne häufiger benutzen (I think that I would like to use this system frequently)

(Q2) Ich finde das System unnötig komplex (I found the system unnecessarily complex)

(Q3) Ich finde das Programm ist einfach zu benutzen (I thought the system was easy to use)

(Q4) Ich denke ich würde die Unterstützung eines Technikers / Informatikers brauchen, um in der Lage zu sein, das Program zu benutzen (I think that I would need the support of a technical person to be able to use this system)

(Q5) Ich finde die verschiedenen Funktionen in diesem Programm sind gut integriert (I found the various functions in this system were well integrated)

(Q6) Ich denke, es gibt zu viele Inkonsistenzen in diesem Programm (I thought there was too much inconsistency in this system)

(Q7) Ich könnte mir vorstellen, dass die meisten Leute sehr schnell lernen würden mit diesem Programm umzugehen (I would imagine that most people would learn to use this system very quickly)

(Q8) Ich fand das Programm sehr umständlich im Gebrauch (I found the system very cumbersome to use)

(Q9) Ich fühlte mich sehr sicher bei der Benutzung des Programmes (I felt very confident using the system)

(Q10) Ich musste eine Menge lernen, bevor ich mit diesem Programm zurecht kam (I needed to learn a lot of things before I could get going with this system)

However, the students had problems with question 6 not understanding what was meant with inconsistency and needed further explanation.

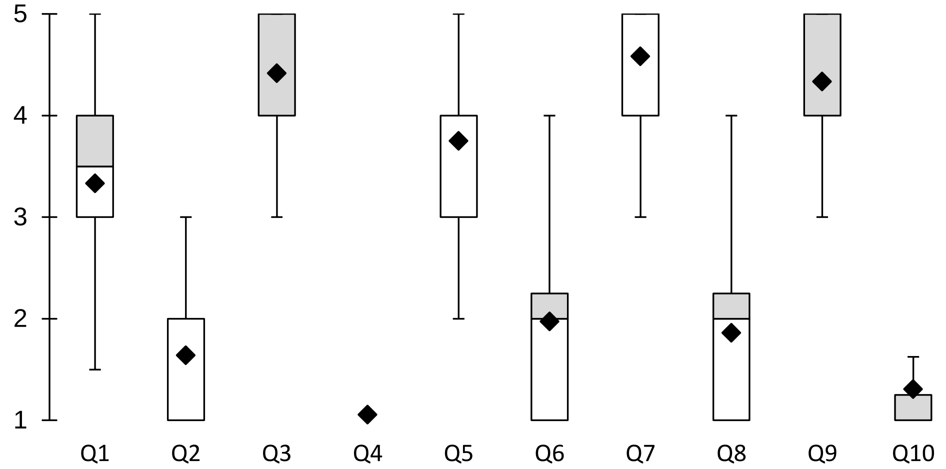


Figure 8.1: Results of the System Usability Scale test

The results shown in Figure 8.1 indicate that the LASAD framework used in this study was perceived as adequate tool to support argumentation. Questions 2, 3, 4 and 10 highlight the ease of use of the system. In combination with the 15 minutes tutorial video this also confirms prior results reported by Van Gelder (2003) indicating that the box-and-arrow / graph visualization is intuitive to most users. Thus, there is no reason to expect a detrimental influence of the system on the outcomes of the study.

Overall Effects on Argumentation Abilities and Domain Knowledge

Based on the scores of the argumentation ability tests ($m(T1) = 1.611$, $sd = 1.02$; $m(T2) = 2.056$, $sd = 0.89$; $m(T3) = 2.083$, $sd = 1.08$; scale ranging from 0 to 4 points), a repeated measures ANOVA was calculated. This showed no statistically significant gains in argumentation skills, but a tendency ($F(2, 66) = 2.907$, $p = 0.062$). The between-subject factor "ontology" did not cause a significant effect ($F(2, 33) = 0.745$, $p = 0.483$).

Regarding the domain knowledge, a significant gain between pre/post-test scores was consistently achieved. In topic 1 (The potential of alternative concepts for automotives), the pre-test resulted in m = 0.92 (sd = 0.77), whereas the post-test resulted in $m = 2.97$ ($sd = 0.88$; based on paired samples t-test: $t(35) = 10.330$, $p < 0.001$; scale ranging from 0 to 4 points). In topic 2 (The German energy mixture in 2030) the pre-test resulted in $m = 2.31$ ($sd = 1.04$), whereas the post-test resulted in $m = 3.42$ ($sd = 0.84$; based on paired samples t-test: $t(35) = -5.976$, $p < 0.001$). Concerning the gain of domain knowledge, there was neither a significant difference between individual/collaborative use of the system nor between the different ontologies.

The Effects of Collaboration on the Argumentation Outcome

An ANOVA highlighted significant differences between individual and collaborative argument maps as shown in Table 2. In comparison, collaborative argument maps contained a larger amount of elements (that is, boxes and relations between them) used overall ($F(1, 46) = 18.954$, $p < 0.001$) and a higher percentage of material used twice ($F(1, 46) = 6.983$, $p = 0.011$). Contrary to the expectations, the percentage of given material used did not differ significantly between individual and collaborative argumentation ($F(1, 46) = 0.932$, $p = 0.339$). Instead, group members provided significantly more own contributions (not derived from given material) ($F(1, 46) = 13.524$, $p < 0.001$) than individual arguers. Hypothesis C5 (in groups the members will review each others work and, hence, groups will make less mistakes compared to individuals), measured by the percentage of wrongly used elements, has to be rejected ($F(1, 46) = 0.956$, $p = 0.333$). In fact, mistakes made in the group phases were often very similar to those made in the individual phases, e.g. wrong directions of relations. Thus, hypothesis C1 (group work $\rightarrow$ higher quality) is only partially supported. To measure the motivation of the participants, the statements in the personal interviews conducted after the study were analyzed. Here, all groups agreed (after short

Description	**Individual** ($n = 36$)	**Collaborative** ($n = 12$)
Overall number of used elements in the workspace	$m = 104.00$ ($sd = 29.72$)	$m = 143.25$ ($sd = 15.80$)
Number of own contributions (not derived from given material)	$m = 9.97$ ($sd = 5.43$)	$m = 18.58$ ($sd = 10.61$)
Percentage of material used twice	$m = 5.03\%$ ($sd = 4.23$)	$m = 9.14\%$ ($sd = 5.83$)
Percentage of erroneous used elements	$m = 22.46\%$ ($sd = 18.80$)	$m = 28.83\%$ ($sd = 21.73$)

Table 8.1: Comparison between individual and collaborative argument maps

discussions) that working in groups was more motivating than working alone (hypothesis C2). This is supported by the observations of the experimenter, who stated that sometimes the participants in the individual sessions made a bored impression, as opposed to the collaborative sessions. Also, the groups always used all the time for their tasks, while some individuals finished early. Among the study participants, the question about the optimal group size for argumentation was discussed controversially. The majority agreed on two to three people arguing together, larger groups and the resulting growing needs for coordination were seen as potentially detrimental for the overall results.

Hypothesis C3 (collaboration $\rightarrow$ participation drop of single users) is not easy to evaluate. Basically, it has to be investigated if users, when working together, became less active. To do so, the proportion of elements of each user in the collaborative sessions was computed first resulting in $min = 0.16$, $max = 0.59$, $m = 0.33$, $sd = 0.12$ - that is, single users created between 16% and 59% of a collaborative map. Apparently, there were thus no "drop-outs" and no dominating users creating the whole map alone. To represent how active a user is in individual sessions (as compared to his or her peers), also the proportion of each user's argument elements was computed in his or her individual session to the sum of elements of all individual maps of his or her group members ($min = 0.19$, $max = 0.51$, $m = 0.33$, $sd = 0.07$). These two values resulted in a significant Pearson correlation of $\rho = 0.428$ ($p = 0.009$). Thus, hypothesis C3 can be rejected: users who are generally (in)active in individual sessions exhibit the same attitude also in collaborative sessions.

The hypothesis that working in groups might lead to a large amount of off-topic talk (hypothesis C4) could not be confirmed, as Table 8.2 shows. In the argument graphs, there were in fact no noteworthy off-topic contributions at all. The chat, embedded in the tool, seems to work quite well to avoid off-topic talk in the map.

Map No.	**Ontology**	**# Content Episodes**	**# Structure & Coordination Episodes**	**# Off-topic Episodes**	**Overall**
1	Simple	6 (33.3%)	12 (66.7%)	0 (0.0%)	18
2	Simple	3 (33.3%)	4 (44.4%)	2 (22.2%)	9
3	Simple	2 (100.0%)	0 (0.0%)	0 (0.0%)	2

Map No.	Ontology	# Content Episodes	# Structure & Coordination Episodes	# Off-topic Episodes	Overall
4	Simple	14 (51.9%)	13 (48.1%)	0 (0.0%)	27
m(Simple)		*6.25* (44.6%)	*7.25* (51.8%)	*0.5* (3.6%)	*14*
5	Toulmin	5 (38.5%)	8 (61.5%)	0 (0.0%)	13
6	Toulmin	4 (11.8%)	21 (61.8%)	9 (26.5%)	34
7	Toulmin	5 (83.3%)	1 (16.7%)	0 (0.0%)	6
8	Toulmin	6 (46.2%)	6 (46.2%)	1 (7.7%)	13
m(Toulmin)		*5.0* (30.3%	*9.0* (54.5%)	*2.5* (15.2%)	*16.5*
9	Specific	6 (31.6%)	13 (68.4%)	0 (0.0%)	19
10	Specific	1 (4.2%)	20 (83.3%)	3 (12.5%)	24
11	Specific	6 (42.9%)	6 (42.9%)	2 (14.3%)	14
12	Specific	6 (37.5%)	8 (50.0%)	2 (12.5%)	16
m(Specific)		*4.75* (26.0%)	*11.75* (64.4%)	*1.75* (9.6%)	*18.25*
m(Overall)		*5.33* (32.8%)	*9.33* (57.4%)	*1.58* (9.7%)	*16.25*

Table 8.2: Overview of average chat episodes per ontology in multi-user maps

The Effects of Ontology on the Argumentation Outcome

Based on the structural assessment of the maps (with respect to starting hypothesis, conclusion and clear grouping), no significant difference between different ontology conditions could be identified and, hence, hypothesis O1 (higher structural degree of ontology → improved structure of the argument) has to be rejected. However, users of the Toulmin-based ontology did show a tendency not to use a starting hypothesis ($F(2, 45) = 3.100$, $p = 0.055$), which is not really surprising as this ontology follows a different model of argumentation (beginning with data and then drawing a conclusion) and there is no explicit hypothesis element in the ontology.

Ontology	Average percentage of wrongly used elements
Simple	$m = 10.25\%$ ($sd = 10.80$)
Toulmin	$m = 41.29\%$ ($sd = 16.48$)
Specific	$m = 20.63\%$ ($sd = 16.57$)

Table 8.3: Overview of wrongly used ontology elements

A difference between ontologies was found in the percentage of wrongly used elements, e.g. using a hypothesis box to represent a fact or to ignore the direction of a pro relation ($F(2, 45) = 18.082$, $p < 0.001$). A post-hoc Tukey HSD test indicated that there was a significantly higher error rate (shown in Table 8.3) in the Toulmin condition than in the others (p < 0.001 for Toulmin vs. Simple and $p < 0.001$ for Toulmin vs. Specific).

Hypothesis O2 (detailed ontology → elaborated arguments) could be confirmed partly. An ANOVA showed no significant differences ($F(2, 45) = 1.909$, $p = 0.160$) between ontologies with respect to the percentage of given material being used. However, a non-parametric Kruskal-Wallis test indicated that the amount of own contributions (not derived from given material) did differ significantly ($p = 0.034$) as shown in Figure 8.2.

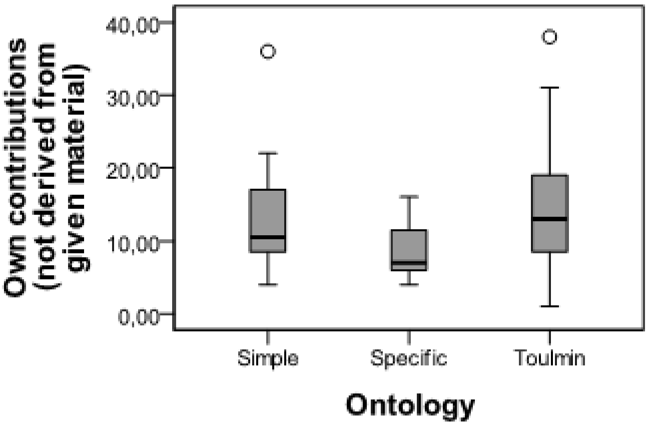


Figure 8.2: Differences of own contributions (not derived from given material) used between ontologies (taken from Loll & Pinkwart, 2011)

Interaction Effects

Regarding hypothesis I1 (ontology will influence the degree of collaboration), the number of relations between elements of different authors in relation to the overall number of links was analyzed as an indicator of the degree of collaboration (since this reflects the inter-relatedness of contributions from different users). An ANOVA did not reveal any significant difference between ontologies ($F(2, 9) = 1.689$, $p = 0.238$). Thus, the hypothesis could not be confirmed. Similarly, the results of a comparison of the number of chat messages used in different ontology conditions did not show any significant differences between ontologies as well (Content episodes: $F(2, 9) = 0.212$, $p = 0.813$; Structure & Coordination episodes: $F(2, 9) = 0.408$, $p = 0.676$; Off-topic episodes: $F(2, 9) = 0.568$, $p = 0.586$).

The comparison of the chat messages can be used for the investigation of hypothesis I2 (highly structured ontology will be detrimental to collaborative argumentation) as well, showing that the amount of needed coordination of structure and activities are not dependent on the complexity of the argument ontology. In addition, there was no significant interaction effect between individual / group argumentation and the ontology (F(2, 42) = 0.605, p = 0.551) in terms of the number of erroneously used elements for argumentation. As such, hypothesis I2 has to be rejected.

8.1.8 Discussion

Regarding the knowledge and the argumentation tests, the results are not surprising: the increase of domain knowledge was an expected side-effect: if students argue about a topic for a longer time with additional material, the result that they have gained knowledge in this field can be expected. However, the motivation to learn by arguing may be increased compared to classic learning. The positive trends shown by the argumentation ability tests is more interesting and needs to be further evaluated in long-term studies – 4 hours use of an argumentation system might not have been enough to come to significant effects at the 0.05 level.

With respect to collaboration, the results of the study confirm the possible benefit of collaboration for learning argumentation and are in line with prior findings (e.g., Janssen et al. (2010); Osborne (2010); Sampson & Clark (2008); Schwarz & Glassner (2007); Schwarz et al. (2000)). Against the hypothesis, groups in the study appeared not to have really checked each other's contributions well, but have argued for or against possible arguments, resulting in more elaborated arguments. This is clearly a point that may be worth future investigations as peer-reviews have shown to be an effective learning strategy (Gehringer, 2001; Cho & Schunn, 2007; Loll & Pinkwart, 2009b) and their inclusion into argumentation system could be fruitful. Based on a scripted approach, a peer-review process could be enforced in argumentation systems. Contrary to the results of Schwarz & Glassner (2007), the influence of structural aids and collaboration on the amount of off-topic talk could not be confirmed in the study. Possibly, the presence of a separate chat window was sufficient to keep the resulting argument map "clean".

Concerning the guiding function of the ontology, the results support Suthers (2003) findings. The use of the Toulmin argumentation scheme did lead to a different style of argumentation: While the Toulmin approach is based on data used to draw a conclusion (without any hypotheses), the other ontologies used in the study employ hypotheses that are then backed up with supporting facts. However, evidence that a domain-specific approach is more beneficial for the overall argument quality than a domain-independent one could not be provided.

In addition, the participants in the study had problems with a highly structured argument ontology, confirming prior findings by Suthers (2003) that a broad range of elements may cause problems for students dealing with it: The Toulmin ontology puts excessive demands on the students due to its complexity. In fact, there were students who denied using the ontology correctly at all and only used the colors of the elements as orientation, for example, using the red *on account of* relation as *contra* and the green *since* relation as *pro*. There was no noteworthy difference between the other two ontologies. Limiting, it should be mentioned that the students were not familiar with any argument ontology before the study and the theoretical argument model of Toulmin was definitely the most complicated one in the study. Also, a less elaborated ontology offers simply fewer possibilities to actually use elements incorrectly.

However, the concrete empirical results were only one purpose of the study. The other one was to provide evidence for the suitability of the proposed framework to conduct more comparable research of open issues in computer-supported argumentation (cf. sub-question 4 in chapter 1.2). Did this study provide such evidence? The answer is clearly *yes*. By means of the configuration mechanisms of LASAD the use of multiple different argumentation tools could be avoided. Instead of using one tool per condition or, at least, one tool per independent variable the LASAD framework could be used for the complete study. The benefits are obvious including the abandonment of multiple trainings of participants on different tools and the absence of unintentionally manipulated factors such as the user interface. Thus, this study provided a first answer to the fourth sub-question of this thesis. However, a single study is not mandatory significant. Therefore, chapter 9 will provide additional evidence that LASAD is capable to deal with typical research scenarios.

8.2 Evaluation of the LASAD Authoring Tool

The study presented in 8.1 indicates that LASAD is capable of dealing with typical requirements of concrete research scenarios. However, another important aspect is that researchers or teachers must also be able to configure LASAD in order to benefit from LASAD's flexibility. Therefore, chapter 7.4 introduced the authoring tool to enable even inexperienced users to configure LASAD to their needs. Here, the authoring tool will be evaluated using a multi-level approach.

8.2.1 Research Questions

In order to answer the question whether inexperienced users are able to configure LASAD to their needs, this evaluation will cover the following three dimensions:

Ease of use: Is the structural guidance provided by the tool high enough to avoid a time-consuming training?

Effectiveness: Are even inexperienced users able to configure the LASAD framework to their needs by means of the authoring tool?

Efficiency: How long does it take to configure the framework by means of the authoring tool? Is there a time benefit of using the authoring tool to configure the framework compared to a manual XML based configuration approach?

8.2.2 Design

To actually measure these dimensions, the evaluation of the authoring tool presented here consisted of three steps:

1. Pilot test with experienced LASAD users
2. Controlled lab study with inexperienced LASAD users
3. Remote test with argumentation experts including teachers, researchers, and developers

In the first step (pilot testing), the authoring tool was given to users that have used the LASAD framework before. They were asked to recreate one of the system configurations they used in the past. After that, they were invited to provide free-form feedback. This first step aimed at evaluating if the functional capabilities of the authoring tool are appropriate to be used in real-world scenarios.

The second step (controlled lab study) should provide evidence whether inexperienced users are able to use the authoring tool to configure the system. Compared to the first session, the students were not aware of the features of LASAD. In addition, most of them have never used an argumentation tool before. Therefore, this was a first test to find out whether a short introduction to the system and its purpose, was enough to actually enable students to work with it in a productive manner.

Whereas students today are usually digital natives (that is, grown up with computer technology), they are only one target group of the authoring tool. The second important group consists of teachers, researchers and developers. Here, one cannot expect familiarity with computer tools (at least in case of teachers and researchers) in any case and, for technology adapted persons, unexpected problems may be revealed. Thus, the final third step was designed as the next iterative step in the ongoing usability improvement process.

8.2.3 Part 1: Pilot test

Method

The pilot test was conducted with two PhD students who have used the LASAD framework before. One of them developed an external analysis and feedback client (see chapter 9.1.3), and the other worked on an integration of the LASAD framework into a learning environment as part of the Metafora project (see chapter 9.2.5). Thus, both were aware of the features provided by the framework and did not require any training on the framework. However, the authoring tool was new to them. Therefore, they were given a manual that explained how to configure the system exemplified by means of a short textual tutorial enriched with screenshots. After a week, both were interviewed to get their feedback.

Results

In the unstructured interview, both testers agreed that the authoring tool was adequate to configure LASAD. The interviewer collected a set of suggestions on how to improve the manual as well as the authoring tool itself. Among the suggestions was the provision of screenshots of all elements that could be used in LASAD as part of the manual. Concerning the authoring tool, one tester proposed to clone ontologies so that one does not start from an empty ontology every time small adjustments were needed to be made. In the next iterative refinement step of the development process of the authoring tool, the feedback was considered and applied. Both testers were able to configure LASAD by means of XML; nevertheless in regular bi-weekly discussions that followed after the test, they reported that they indeed prefer to use the authoring tool instead of direct configuration of LASAD because it is less error-prone and, hence, more time efficient.

8.2.4 Part 2: Test with students

Method

The second test with graduate and PhD students was conducted after the pilot test was completed and the feedback of the pilot testers was applied to the manual and the authoring tool. Overall, 10 students were asked personally and participated voluntarily. Even though this is a rather low number, it was expected to be adequate since the literature reports 3-4 people to be enough to cover the most important usability lacks (Krug, 2006, 2010), which was the main purpose of this study step. All of the participants were experienced with computer systems in general and enrolled in either a computer science or business information systems study course. Their ages ranged from 24 to 30 ($m = 26.4$, $sd = 2.01$) years. Each student received a short 5 minute introduction to the LASAD system, the system's goals as well as the purpose of the study from the same tutor. After that, they were asked to work through a hands on tutorial example which is part of the manual of the authoring tool. Questions were answered by the tutor.

Once training was completed, the participants received two tasks in succession. The tasks can be found in Appendix F. Both tasks were designed to be of a similar level of difficulty with respect to the required configuration steps of LASAD. Whereas the first task was formulated to fit to the linear steps of the authoring tool, the second was formulated to be less structured as it would be in reality. To minimize the influence of training effects the tasks were counterbalanced, that is, half of the participants started with the first task whereas the other half started with the second task. The configurations that needed to be done during the studies were designed to fit to existing argumentation systems. Whereas the first task made use of a LARGO style argumentation, the second one used a Belvedere style of argumentation. However, the configurations were modified to use each component of the authoring tool and to be less time consuming, e.g., a short given text was used instead of a transcript with more than 100 lines or some types of relations were dropped because adding them was much the same as for the first one.

Instead of requiring the participants to read the manual before using the authoring tool, they received it as a handout that could be used during the study. The tutor did not answer any questions concerning the authoring tool after the tutorial. Thus, the participants had to consult the manual if they ran into problems.

Finally, all participants were asked for feedback concerning problems and possible improvements with respect to the authoring tool in an open interview with the tutor.

To answer the research questions, all actions of the students were recorded by means of a screen capturing tool. This way, it was possible to stop the required time and identify possible confusions and problems in particular steps. Further, the recordings can be used to identify improvement potentials for future usability oriented changes.

Results

All students were able to solve the tasks they were given without any additional help from the tutor. The concrete results are shown in Table 8.4.

User	**Order**	**Time$_{T1}$ in s**	**Errors$_{T1}$**	**Time$_{T2}$ in s**	**Errors$_{T2}$**
Student 1	T1-T2	648	0	962	0
Student 2	T2-T1	433	0	845	1
Student 3	T1-T2	1046	0	876	2
Student 4	T2-T1	542	0	1123	1
Student 5	T1-T2	688	0	981	1
Student 6	T2-T1	477	1	920	3
Student 7	T1-T2	584	2	648	1
Student 8	T2-T1	546	1	1095	2
Student 9	T1-T2	839	0	891	0
Student 10	T2-T1	585	0	1175	1
m	*T1-T2*	*761*	*0.40*	*871.60*	*0.80*
sd		*184.89*	*0.89*	*132.79*	*0.84*
m	*T2-T1*	*516.60*	*0.40*	*1031.60*	*1.6*
sd		*60.71*	*0.55*	*141.61*	*0.89*
m		*638.80*	*0.40*	*951.60*	*1.20*
sd		*182.82*	*0.70*	*154.47*	*0.92*

Table 8.4: Times and error rates of the authoring tool test

With respect to the efficiency, the required time and the error rate had to be considered. The average time required by the participants for task 1, the structured one, was around ten and a half minutes ($m = 638.8$, $sd = 182.82$ in seconds), whereas the time was around sixteen minutes ($m = 951.6$, $sd = 154.47$ in seconds) for task 2, that is, the unstructured one. Both time frames are considered reasonable, especially since a direct

XML configuration would not be faster according to developer experiences. The error counts for both tasks were low ($m(t_1) = 0.4$, $sd(t_1) = 0.7$; $m(t_2) = 1.2$, $sd(t_2) = 0.7$). In addition, all participants were asked to resolve their mistakes after they reported to be done with the task. Here, each participant was able to fix the problems. In most cases, the students reported simply reading over the relevant passages, especially in the unstructured condition.

The screen recordings in connection with the final interviews with the participants highlighted possible improvements and usability issues. Based on these results the authoring tool was improved in the next iterative development step. The most important change was the addition of a possibility to modify nodes, relations and elements once they have been created.

An in-depth analysis of the data revealed that the users tend to use less time for their second task as Figure 8.3 shows. Thus, one can conclude that additional training would have lowered the time required to solve the tasks further.

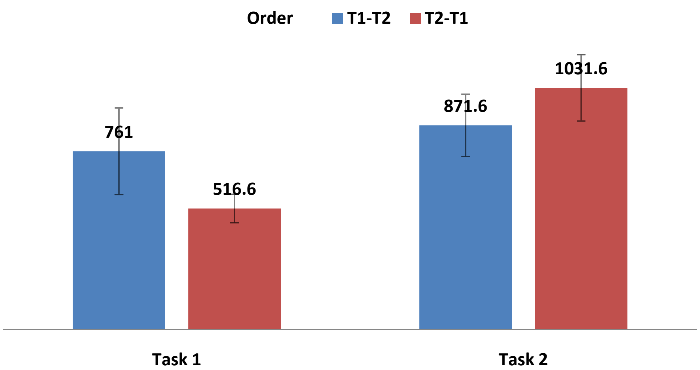


Figure 8.3: Required time (in seconds) on tasks dependent on the task order

Overall, the authoring tool has proven to be a suitable tool to configure LASAD in this study. However, a limiting factor that has to be mentioned is that all participants were students with a background in computer science or business information systems. Thus, they are used to typical problems that may occur using computers.

8.2.5 Part 3: Test with researchers, teachers, and developers

Method

The final test with argumentation experts was done via the web. Here, researchers, teachers and developers of argumentation systems were invited to participate voluntarily in a public testing of the authoring tool. The invitation was sent personally via e-mail to the experts that have either pilot tested or answered the survey presented in section 4. However, survey participants who reported themselves as not being experts were dropped. Overall, 100 experts were invited to test the authoring tool. As motivation to provide feedback in form of an online questionnaire (presented in Appendix F.3) an Apple iPod Nano was raffled among all participants. None of the experts were familiar with the LASAD framework. They only received a short paragraph explaining the goals of LASAD as well as a reference to the projects homepage. By means of an online demo version of the system, the experts were able to test the framework. In addition, they were provided with the improved manual of the prior tests with the authoring tool. In particular they were asked to go through the example presented in one of the sections of the manual. In case of problems, they were invited to ask questions via e-mail.

Results

In total, 19 experts provided feedback. Unfortunately not all of them used the authoring tool of the system. Instead, some experts used the LASAD argumentation tool only. The distinction between those who used the authoring tool and those who did not is unclear, because some of the responses were provided anonymously and, thus, could not be tracked in the logs. Therefore, the only indicator whether the anonymous respondents used the authoring tool comes from the open-ended comments. Even so, not all of them provided those. Consequently, the Likert-scale questions could not be used for the analysis. Further, verifiably none of the experts went through the example. This has major impacts on the results. Since nobody received any training, the concrete use of the authoring tool was significantly harder.

However, the open-ended questions are of higher interest anyway. Based on the experts' comments, there were two main problems: On the one hand, some of the experts reported problems with hardly readable instructions caused by limited screen space. On the other hand, the experts bemoan the terminology used in the authoring tool. Quotes that back these point up are the following:

> *"The instructions did not display correctly on my monitor - they were partly hidden by another part of the interface."*

> *"In my opinion the word ontology is not the most appropriate one in this settings, as its meaning for common public outside argumentation community is different. This misunderstanding should cause some affordance problems to common users."*

> *"I did not read the manual before trying to author. I probably should have done that. I was a little confused by the concepts of template, session, etc. the first time I used the author tool. An experienced user would probably not be confused by that."*

In addition to the problems mentioned by the experts, there were proposals for improvement given. These proposals, however, focused on the general LASAD framework as well as on the authoring tool. With respect to the authoring tool, the experts asked for concrete use cases that could serve as examples. One of the experts mentioned that the manual was too lengthy. This explains why none of the experts went through the example of the manual. However, in connection with the terminology problems described previously, it is also clear that some kind of training is indeed required to understand the underlying concepts of LASAD and the possibilities of the authoring tool.

On the functional side, one of the experts proposed a direct link from the last step of the authoring tool to the argument map that has been created. Another one proposed to allow for on-the-fly changes of the ontology. However, these proposals are somewhat contradictory to the concept of the authoring tool. As described in chapter 7.4, the authoring tools' purpose is to define the argument model and interface elements beforehand. That is, the users were not intended to modify the ontology, but use it. Thus, the restrictions that are implied by the system configurations are on purpose. A small functional extension proposed by another expert is the possibility to rename ontologies, templates and sessions after they have been created. In addition, one expert mentioned that he was not able to emulate his system (Carneades), because the authoring tool did not include the possibility to configure artificial intelligence rules and agents.

In addition, there were possible improvements for the non-authoring part of the LASAD framework. Here, the printing of argument maps as well as the possibility to hide link panels via a button was mentioned. These features, in particular the first one, are likely to be integrated in a future version of LASAD.

Nevertheless, about a third of the participants did not report any problems. In addition, there were experts who explicitly asked to use LASAD for their purposes because they highly appreciated the flexibility of the framework in combination with the authoring tool.

8.2.6 Discussion & Outlook

The authoring tool was designed to support people in configuring LASAD. The key goal was to avoid the writing of pure XML code for configuration to enable even non-programmers to benefit from the flexibility mechanisms of LASAD. In order to evaluate the adequacy of the authoring tool to fulfill these needs, three key values have been evaluated: Ease of use, effectiveness, and efficiency.

The first point, that is, the ease of use, have shown to be improvable. On the one hand, there have been proposals given to extend the functionality of the authoring tool. This way, the usability could be improved, and complicated or time consuming workarounds could be avoided. An example is the clone feature for ontologies, which allows to easily modify a single factor without a complete rework of the rest. On the other hand, it turned out that training is still required. Whereas the students who were required to go through a short hands-on tutorial did not report any problems the experts, who did not go through the tutorial, did. As correctly mentioned by one of the experts, "*it's rare that folks will read manuals*" especially a rather lengthy one. However, parts of the manual are prerequisite to understand the underlying concept of the configuration mechanisms in LASAD. In particular, the underlying terminology (consisting of ontology, template and session) is not obviously clear and was the main reason for problems. Thus, future versions of the authoring tool will move away from this terminology in favor of a more intuitive one such as *underlying argument model* (instead of ontology), *argumentation context* (instead of template) and *argumentation session*. Additionally, important parts of the manual could be replaced by a video tutorial which is more likely to be viewed. In order to solve the screen space issue a provision of *"more context-sensitive instructions"* may be helpful in the experts' opinions.

The second point, that is, the effectiveness, have shown to be good. The fact that experienced LASAD users such as those who were participating in the pilot test (see section 8.2.3) favor the authoring tool for configuring LASAD over the XML approach clearly highlights that it is an effective tool.

The third point, that is, the efficiency, is largely confirmed by the test with students (see section 8.2.4). The required periods of time to configure the system are rather short and the results indicated that they can be reduced further once the users are familiar with the tool. However, due to the limited number of participants this can only be understood as indicator without statistical backup.

On the functional side, missing features (as reported by the experts) may be added later when the authoring tool will be extended to cover scripting and artificial intelligence parts as well.

In conclusion, the authoring tool approach to facilitate the basic configuration of LASAD was successful. The next steps will concentrate on a further simplification in order to achieve an improved user experienced. Additionally, future tests can now be compared to the current state of the authoring tool and, thus, be more comparable.

9 External Applications of the LASAD Framework

Even though chapter 8 provided evidence that the framework is able to be used in developer controlled situations, this is only one side of the coin. As a matter of fact, in real world scenarios additional problems may occur that have not been revealed during lab studies. Therefore, an application to concrete practical problems outside the author's range is essential. At the same time such an approach will show both the software's generality and its maturity to be used in real world scenarios.

As mentioned in the previous chapters, there are multiple target groups of the proposed framework. On the one hand, teachers and students who will use the framework in education. On the other hand, researchers and developers who are eager to use the tool for their purposes, for instance to evaluate the influence of single factors on the argumentation outcomes or to simplify the development of systems that can be used in learning or researching contexts. Therefore, it is important to highlight the successful application of the framework to concrete situation in practice. In this chapter, I would like to point to work of other developers, researchers, and teachers who were using the LASAD framework. Therefore, I will sum up the work done and highlight the use of LASAD in these contexts.

As a limiting factor, it has to be mentioned that all of the work presented here was supported, at least in parts, by the author. For each application, which will be described in this chapter, the concrete role of the author will be mentioned at the end. However, a completely independent use of LASAD will be possible in future. Therefore, the reference implementation of the framework will be made available under an open source license, when this thesis will be in press.

9.1 Developments

Concerning the developer's view on the LASAD framework, there has been efforts to extend the framework mostly on the client layer. As mentioned in chapter 6.1, basically three kinds of extensions can be applied here: (1) additional views, (2) additional clients or (3) additional argumentation and interface elements.

9.1.1 Emulating Existing Argumentation Frameworks

One of the key motivations behind LASAD is the provision of increased comparability of research results. In order to achieve this comparability, it is essential that LASAD can actually be applied to multiple domains. One of the first steps, which can be seen as proof of concept has been done by Brenner (2011). In his diploma thesis, he used the LASAD framework to emulate the existing argumentation system LARGO. A direct comparison between LARGO and the LASAD emulation of it is shown in Figure 9.1. As mentioned in chapter 3, LARGO has been designed to support law students in acquiring argumentation skills. Brenner evaluated the available elements present in a very early prototype of LASAD and extended the framework in order to be able to deal with the requirements present in the application area of LARGO, that is, legal argumentation with focus on hypothetical reasoning.

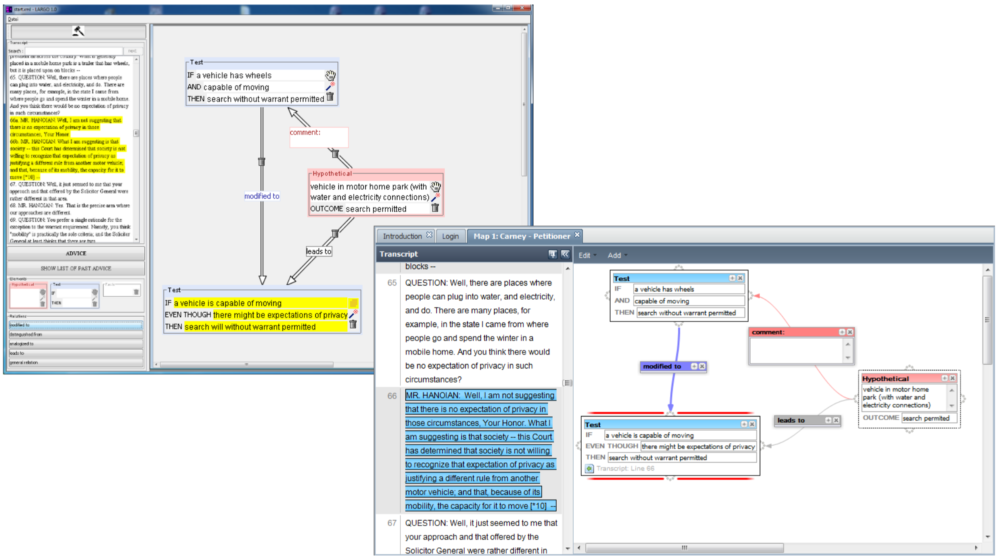


Figure 9.1: Emulation of the LARGO system in LASAD

In order to evaluate the success of the emulation, he conducted a lab study with 11 paid participants at the University of Pittsburgh in spring 2010. In the focus of the study was the usability of the system. His results showed that LASAD is basically able to emulate LARGO. The emulation achieved an average score of 65.9 out of 100 (Brenner, 2011) on the System Usability Scale (SUS) (Brooke, 1996), which is equivalent to a "D" grade according to Bangor et al. (2008, 2009). The overall score seems rather low at first glance. However, the system was in a very early stage and technical problems occurred frequently. On the feature level, however, the system was able to compete with LARGO's basic functionalities and was even superior when considering the collaborative argument modeling functions that were not present in LARGO at all. As described in chapter 8.1.7, a later study with a matured version of LASAD showed in increased mean SUS score of 81.46, which is equivalent to a "B" grade.

This work has been done as part of a diploma thesis, supervised by the author. The development and evaluation has been done independently.

A similar proof of concept by means of tool emulation is described in (Loll et al., 2011b). In addition to the LARGO emulation, there is also a successful emulation of the well-known Belvedere system described. The latter rebuilt the basic functionalities of Belvedere, that is, the ontology of the latest version (version 4.1) in combination with collaboration support of older versions. However, the feedback component of Belvedere was left out. Nevertheless, the LASAD framework would allow to emulate this by means of an additional feedback client as well.

9.1.2 A Table-Based Visualization

Whereas the work of Brenner (2011) concentrated on the extension of the framework via additional argumentation and interface elements, Zhao (2011) chose a different way for extension. He enriched the visual capabilities of LASAD by a table view (see Figure 9.2). In particular, he addressed the problem of using different visualizations on the same data basis. Concrete issues that were addressed in this work include the missing meta data of elements that has been created in the table view (as they do not need any (x, y) coordinates for correct positioning of elements) as well as the synchronization of cursor-movements of users that are using a different zoom level.

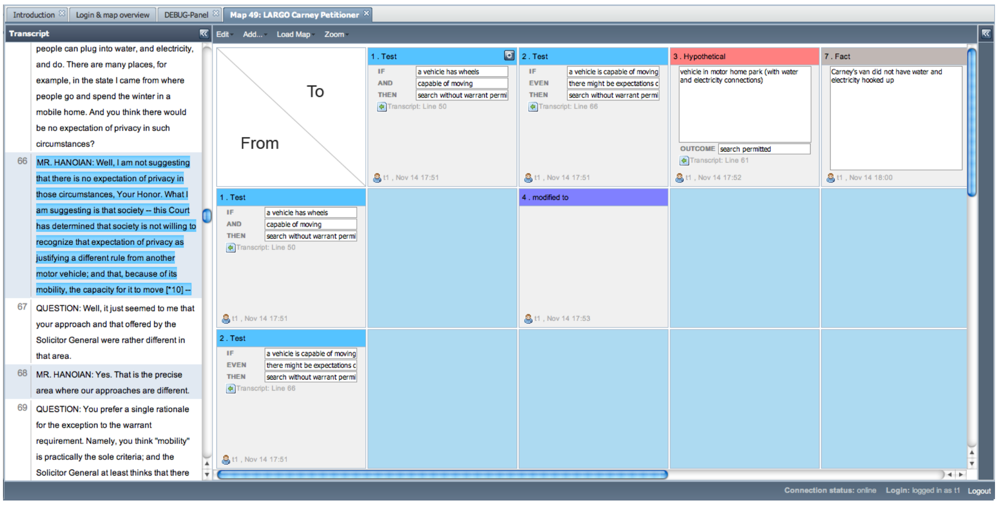


Figure 9.2: Alternative visualization of argument structures in LASAD - Table view

This work has been done as part of a diploma thesis, supervised by the author. The development has been done independently in most parts. However, the author provided guidance on the code level.

9.1.3 Additional Clients

A third example of extensions for LASAD is the work of (Scheuer et al., 2009) and (Neu, 2010), who created additional clients for the LASAD framework.

An Analysis & Feedback Client

The first external client, developed by (Scheuer et al., 2009), was a Java based artificial intelligence agent. The agent applied existing methods of artificial intelligence former used in the context of LARGO (Pinkwart et al., 2006, 2007, 2008) and Digalo / ARGUNAUT (McLaren et al., 2010) to the emulation of these systems in LASAD. Currently, this analysis and feedback agent is extended to provide domain-independent support that is independent from the underlying ontology.

The authors of the analysis and feedback client were part of the LASAD project team. Here, the author's role was to provide guidance on how to integrate the analysis and feedback client with the reference implementation of LASAD. However, the concrete implementation was done completely independent.

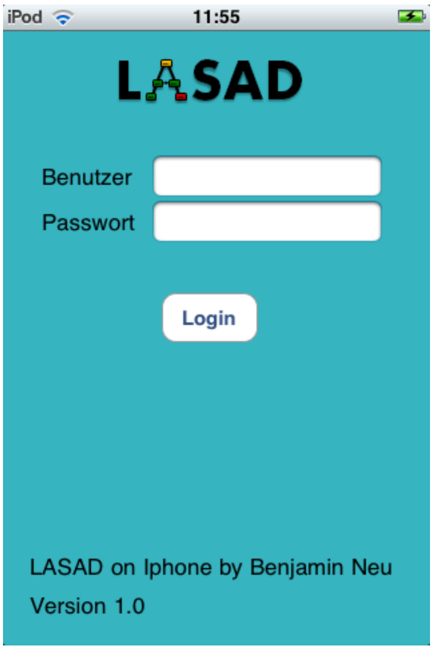
A Client for Mobile Devices

The second external client, developed by Neu (2010), is an iOS[1] based client which intended to transfer the framework's client to the growing field of mobile computing. By means of this client, he investigated the question how to deal with limitations concerning the screen size of typical mobile devices. The iOS prototype is shown in Figure 9.3.

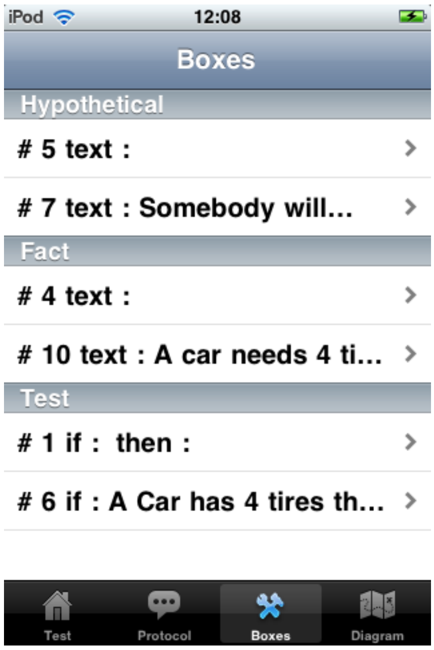
This work has been done as part of a bachelor thesis, supervised by the author. The development has been done independently. However, the author provided high-level guidance on the integration of the iOS client with the reference implementation of the framework.

With respect to the LASAD framework, the developments of Scheuer et al. (2009) and Neu (2010) highlight the platform independency of the approach, not only with respect to the operating system, but also with respect to the underlying hardware. In addition, they show how flexible the framework and the underlying data format is.

[1] iOS is the operating system that is used on most Apple devices including the iPhone, iPod, and iPad



(a) Login screen



(b) Overview of argument session elements

Figure 9.3: LASAD - iOS Client

9.2 Research & Teaching

Whereas the extensions of the framework provide evidence for the flexibility of LASAD, another crucial question is still open: Is it suitable to be used for concrete research tasks and can it be successfully applied in learning scenarios?

9.2.1 The Diagnosticity of Argument Diagrams

In his PhD thesis Lynch (2012)[2] investigated the question whether diagrams as modeled in argumentation systems such as LASAD can be used to diagnose argumentation. More concrete, he investigated the following hypotheses (directly taken from the proposal of his PhD thesis):

H1. The use of diagrammatic models of argument to examine or annotate existing arguments will improve novices' ability to make novel arguments.

[2]The PhD thesis is not finished yet. The numbers presented here are based on the dissertation proposal and, hence, may not be final.

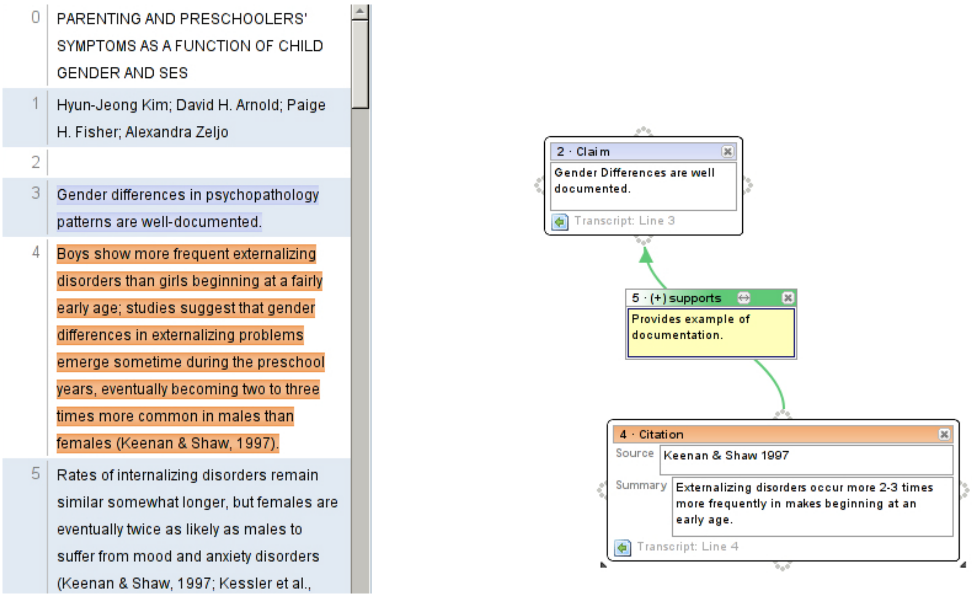


Figure 9.4: LASAD annotation example (taken from Lynch (2011))

H2. Student-produced argument diagrams (involving both annotation of existing arguments and planning of new arguments) can be used to diagnose students' performance.

This second hypothesis can be split up into the following subhypotheses:

a) Student performance may be assessed by *human* graders.

b) Student performance may be assessed through *automated* diagram analysis.

To evaluate the hypotheses, he used data from the LARGO system (Pinkwart et al., 2006, 2007, 2008) as well as data from an early prototype version of LASAD. The latter was intended to be used in three studies.

In the first study, LASAD was used to diagram and plan research reports in four series of psychology research methods and biological writing courses at the University of Pittsburgh as well as the Carnegie Mellon University. The first course — writing for biology majors — at the University of Pittsburgh consisted of 21 students who were required to perform at least one annotation task, develop one planning diagram and review two papers from peers. An example of the annotation task is given in Figure 9.4. This resulted in a total of 84 diagrams. The second and third course was held in form of psychology courses at the Carnegie Mellon University. The courses consisted of 19 and 21 students respectively. Here, each student was expected to annotate one prior paper and produce one planning diagram. Thus, there were 80 diagrams in total. The final course on research methods was held at the University of Pittsburgh again.

It consisted of nine sections of roughly 20 students each. In seven of these sections, the LASAD system was used. The students were either working individually or in pairs on the planning and diagram tasks. None of the sections produced review diagrams. Overall, 224 diagrams in total have been created.

In the second study, 51 students of a legal writing course were intended to be using LASAD in order to create outlines of written arguments taken from U.S. court cases. However, due to technical problems concerning the limited capacity of the wireless network, the study was not completed.

The third study was conducted in the constitutional law course at the School of Law of the University of Pittsburgh. Here, the pre-configured LARGO configuration of LASAD was used without any additional help. The study involved 63 students. Each students created four diagrams covering three U.S. Surpreme Court oral arguments, that is, there were 252 diagrams in total. The tasks were performed from home to circumvent limitations of the network connection.

While writing these lines, the evaluation of the results is not yet finished. However, resulting from the experiences with LASAD, Lynch compiled a list of possible improvements for future developments. These proposals included — on the functional side — the possibility to zoom in and out of diagrams as well as a helping mechanism for transitioning from diagrams to a more linear structure as present in textual reports. The former can be done by means of browser functionality. Here, each web browser is able to zoom in or out of web pages. However, an additional overview in form of a mini map (cf. chapter 3.3) is imaginable for future development. The latter, that is, a textual outline of the argument structure is not present in LASAD. An integration into LASAD can be done by means of a new visualization.

On the non-functional side of the proposals, the performance of LASAD in combination with a rather high network traffic was denunciated. However, when he used LASAD it was in an early stage of the development process, which was not intended to be used in concrete research scenarios. In addition, the server was located in Germany, whereas the study took place in the United States which was detrimental to the overall response time. Nevertheless, the results of Lynch's test led to a rework of the server of LASAD. Instead of using Enterprise Java Beans, which resulted in a lot of overhead, the server was migrated to plain Java. Since most of the reported problems occurred when using the system with multiple users at the same time, the degree of parallelization was increased. As shown in later studies (cf. sections 9.2.2 and 9.2.3), the changes were fruitful and the problems did not occur again.

This work has been done independently of the author who only answered (mainly technical) questions to the LASAD framework and it's configuration. In contrast to the other studies, the LASAD server in this study ran Linux instead of Windows which was used for most of the other studies.

9.2.2 Enhancing Discussions in Social Networks with Argument Graphs

Lately, social networks such as Facebook, SchülerVZ/StudiVZ/MeinVZ or Google+ became increasingly popular. One of their main purposes is to foster communication between users. However, most discussions conducted in these environments are superficial and lack in-depth reasoning. An imaginable way to promote the elaboration of arguments is the parallel use of argumentation systems. Therefore, a study with 20 psychology and educational science students was conducted at Saarland University in summer 2011 by Oliver Scheuer and colleagues. The main hypothesis was that the use of LASAD in connection with Facebook would lead to improved arguments within the Facebook discussions. As a direct consequence, there was a higher learning benefit with respect to the discussed topic compared to plain text discussions in Facebook without the use of LASAD hypothesized.

In order to test this hypothesis, the study followed a between subjects design with two conditions. The first condition included the use of LASAD, whereas the second did not. An example of an LASAD diagram constructed within the tool condition is given in Figure 9.5. In both conditions, there were student dyads. These dyads were arranged to maximize opinion differences in order to motivate discussion. The discussion topic was to evaluate the pros and cons of behavioral teaching methods. The concrete activities of the study consisted of the following steps:

1. Pre-test
2. Reading of preparatory texts about behaviorism (15 minutes)
3. Preparation of individual statement about behaviorism (5 minutes, used to assign dyads)
4. LASAD training (2 minutes in LASAD condition)
5. Individual argument creation in LASAD (15 minutes in LASAD condition)
6. Facebook discussion in dyads with the goal to agree on a shared conclusion (25 minutes in LASAD condition, 42 in control condition)
7. Post-test

The evaluation of the collected data is currently ongoing. There were no technical problems with LASAD.

This work has been done by members of the LASAD project team. However, the author was neither involved in the study design nor the execution. The author's only contribution was to answer questions concerning the configurability of the reference implementation.

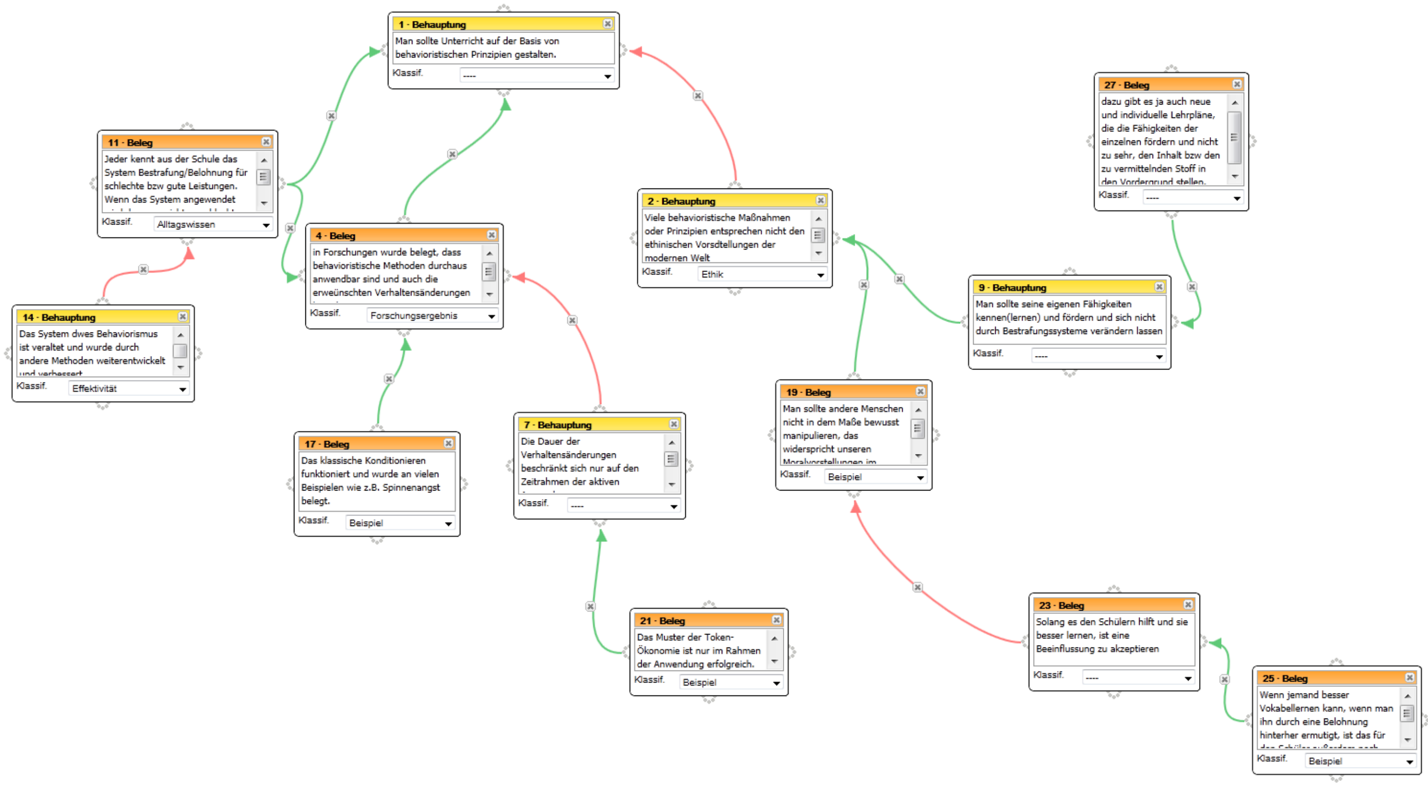


Figure 9.5: Example of a preparation diagram in tool condition

9.2.3 Deepening Argumentative Discourse via Scripted Roles

When arguing about controversial topics a common problem of students is that they do not think in-depth about arguments that are against their own position. Instead, they either agree to easily or they ignore others completely. However, one imaginable way to deal with this problem is the assignment of roles in student discussions. As mentioned in chapter 3.5, the assignment of roles may promote argumentation quality. In order to evaluate this point further, Oliver Scheuer and colleagues conducted a field study in a humanities and social science classroom in summer 2011. Their main hypothesis was that the use of role scripting in LASAD would lead to more in-depth elaboration in form of an increased amount of (counter)arguments, questions, explanations, examples, etc. As direct consequence, they expected a higher gain in domain knowledge than without the role script. The role script consisted of two roles: an interceder and a critic. These roles were connected to a given set of sentence-openers (see chapter 3.3) in the chat which are typical for their role. A snapshot of the argumentation session showing the sentence opener chat interface is presented shown in Figure 9.6.

Overall, 44 students participated. The students were grouped into dyads. While 12 dyads were working with a role script, 10 dyads did not. In each dyad the students were given different background texts containing either pro or con arguments with respect to a given hypothesis about global climate change. The concrete study design involved the following phases:

1. Pre-test (25 minutes)
2. Individual familiarization with LASAD (22 minutes)
3. Collaborative work on task with LASAD
    - Phase 1: Read text and map arguments into LASAD (25 minutes)
    - Phase 2: Discussion about text and extension of main arguments by means of additional points with / without sentence-openers (25 minutes)
    - Phase 3: Discussion between both texts with / without sentence-openers (20 minutes)
    - Phase 4: Discussion of opinion and agreement on a reasoned position with / without sentence-openers (20 minutes)
4. Post-test (25 minutes)

The evaluation of the collected data is currently ongoing. There were no technical problems with LASAD.

This work has been done by members of the LASAD project team. However, the author was neither involved in the study design nor the execution. The author's only contribution was to answer technical questions concerning the extension of the chat to make use of sentence openers.

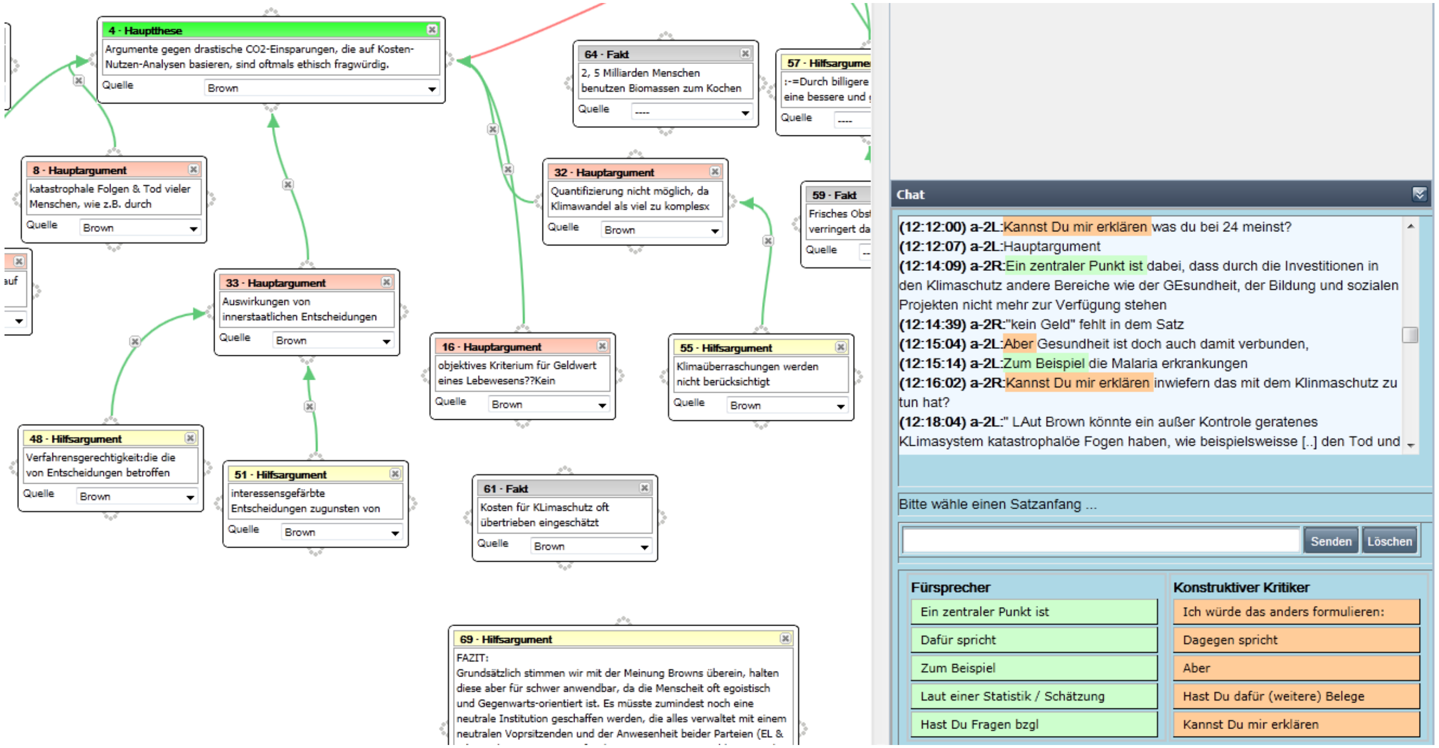


Figure 9.6: Snapshot of argumentation session with scripted roles and sentence openers

9.2.4 Lecture “Computer Supported Collaborative Work”

In the summer term 2011, there was a course in “Computer Supported Collaborative Work” at Clausthal University of Technology held by Prof. Niels Pinkwart. Here, the participants were required to design a lab study in order to pass the exam. One of the student groups designed a study about logical puzzle solving. In this study, they compared grids on paper with the graph-based view of LASAD to aid the solving process. A part of one of the sample solutions for the tasks is shown in Figure 9.7.

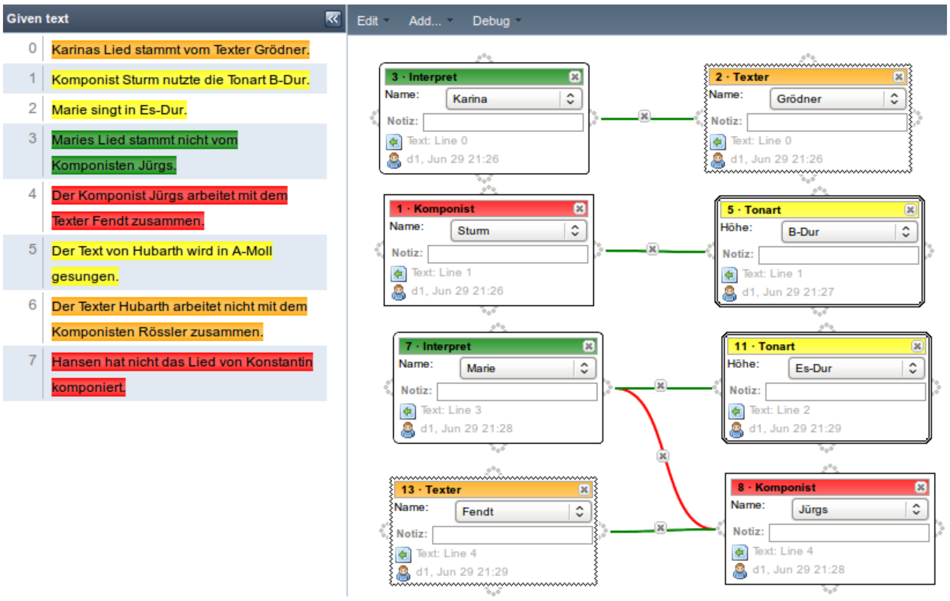


Figure 9.7: Part of a sample solution for a logical puzzle modeled in LASAD. The relations are used to symbolize correct and wrong relations between composer [Komponist], song writer [Texter], interpreter [Interpret], tonality [Tonart] based on the given information shown on the left-hand side.

Even though the participants were not able to solve the puzzle in any of the conditions, the feedback of the students indicated that the graph-based view of LASAD is inferior compared to a paper and pencil table mainly caused by overview reasons. Whereas the results are not surprising, the use of LASAD in the course highlights two important points. On the one hand, it shows that the configurability of LASAD (supported by the authoring tool which was used in this scenario to configure the system) is general enough to enable users to use it in a completely different context (logical puzzle solving instead of argumentation). On the other hand, it shows that the system is usable as supporting tool in teaching.

This instructor of the course was part of the LASAD project team. However, the configuration and the actual use of LASAD has been done independently by the students.

The author's role was to provide a runnable instance of the reference implementation.

9.2.5 The Metafora Project

In addition to the conducted studies, the LASAD framework will be used as part of the Metafora project[3]. The goal of this three-year EU-funded research project, which started in July 2010, is "*the creation of a Computer-Supported Collaborative Learning (CSCL) system to enable 12 to 16-years-old students to learn science and mathematics in an effective and enjoyable way*" (Schwarz et al., 2010, p. 2). Here, Metafora follows an integrated approach, that is, the combination of tools plays an important role. More concrete, the learning process in Metafora is scripted by means of a planning tool.

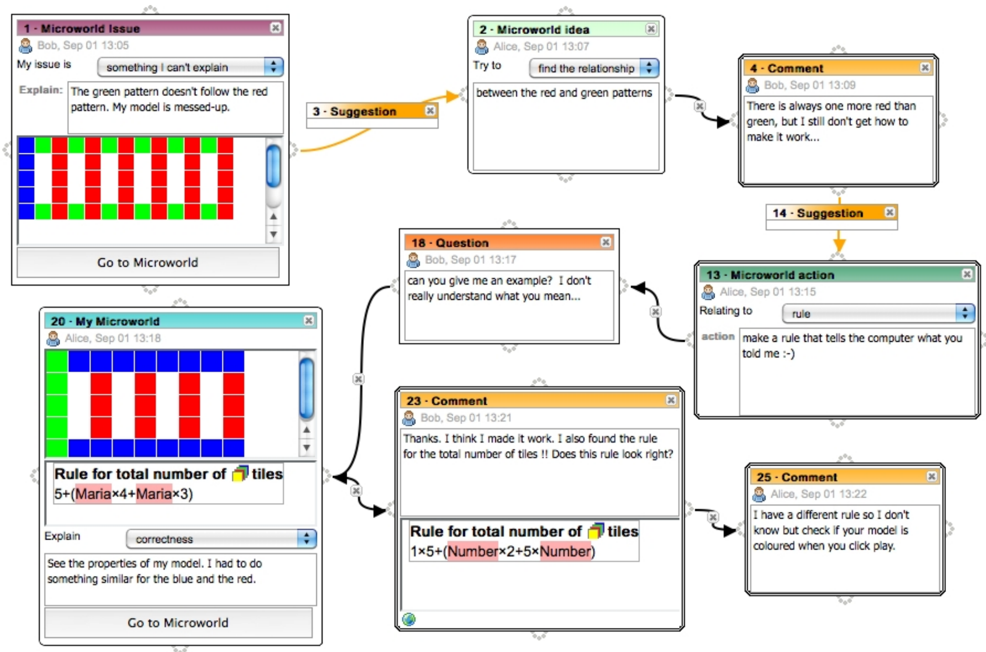


Figure 9.8: Integration of the eXpresser microworld into LASAD (taken from Dragon et al., 2011a)

In this tool multiple phases of learning are connected to activities and roles. Each activity, however, is connected to one or more tools that can be started directly from the plan created beforehand. By means of the collected data from these multiple tools, which may include microworlds, simulations as well as LASAD as discussion environment, concrete feedbacks can be generated. Therefore, each component will forwards

[3]http://www.metafora-project.org

notifications about significant steps in the learning process, so called landmarks, to the Metafora system. Based on the information available, the latter will compile feedbacks for teachers or student to promote learning activities. An imaginable scenario which exemplifies the combination of a microworld designed to learn algebraic rules and LASAD as discussion tool is presented in (Dragon et al., 2011a,b). Here, a first draft of a solution to an algebraic task is used as landmark. When working in dyads, LASAD could be used to discuss alternative solutions once the microworld communicates to the other components that both users completed their draft. After that, LASAD would be able to notify a teacher via the Metafora system when there are not resolvable conflicts. An example of such an integration can be found in Figure 9.8.

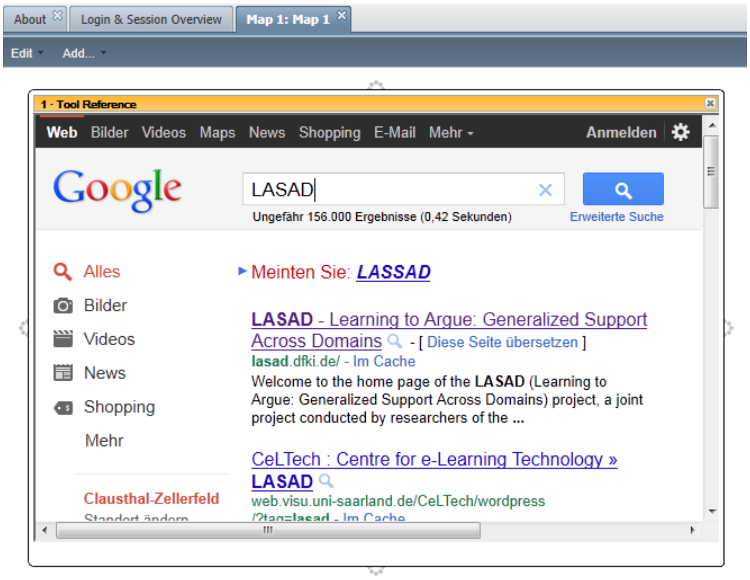


Figure 9.9: Integration of external web pages or web applications into an LASAD ontology element

However, since LASAD is a web-based tool, the integration can be done either by means of a screenshot or by means of an inline web page presented in a LASAD ontology element, which is shown in Figure 9.9. This way a combination of LASAD with a large set of web tools can be realized.

Parts of the LASAD project team are part of the Metafora project team as well. However, the author is not. Thus, the author was only available for technical questions in the starting phase of the project.

9.2.6 LASAD Integration with SWoRD

Another project which will be using LASAD is going to start in fall 2011. Here, the educational peer review tool SWoRD (Cho & Schunn, 2007) will be used in connection with LASAD as diagramming tool. Guided by artificial intelligence techniques, the National Science Foundation (NSF[4]) funded project aims at supporting high school, undergraduate and graduate students in learning scientific writing. More concrete, the process illustrated in Figure 9.10 will be supported.

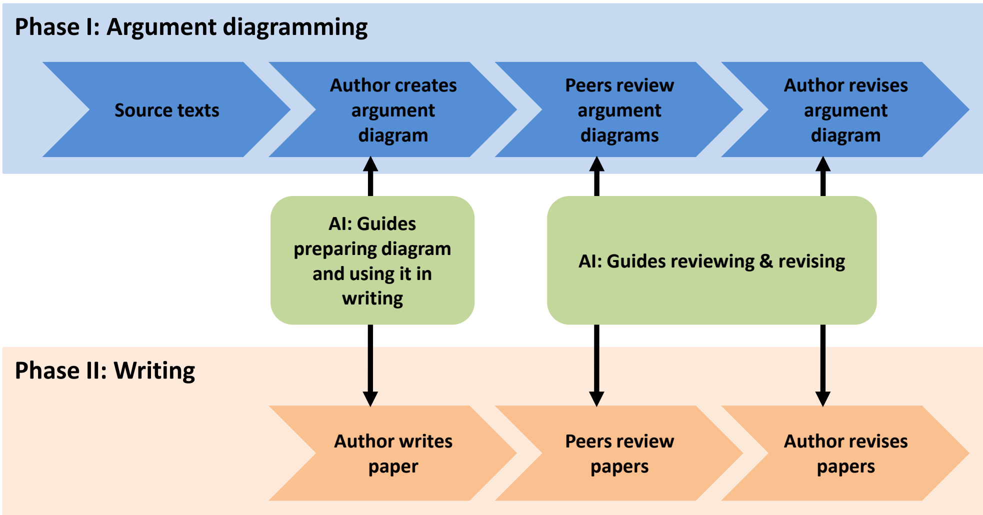


Figure 9.10: Supported problem-solving process

In this project, LASAD will be used to support the peer review process. This support is done via visualization as well as via artificial intelligence feedback.

While writing these lines, the project is still in the planning phase. The author will not be involved in the project. For the proposal phase, the author answered questions concerning the technology side of LASAD, that is, providing advice concerning the potential integration of SWoRD and LASAD.

9.3 Summary & Conclusion

In this chapter, various research, development and teaching contexts involving LASAD have been presented. A summary can be found in Table 9.1.

[4] http://www.nsf.gov

Section	Description	Research / Teaching Focus	Used LASAD Features
9.1.1	Emulating LARGO in LASAD.	Exploration of LASAD's capabilities in domain-specific contexts.	Configuration mechanisms
9.1.2	Development of a table-based visualization as part of the LASAD reference implementation.	Investigation of challenges of mapping different visualizations.	Flexible data format; possibility to add additional visualizations
9.1.3	Extending LASAD by means of an external analysis and feedback client.	Applying existing feedback technology from ARGUNAUT and LARGO to LASAD; building a general feedback infrastructure for future extensions.	Public interfaces of LASAD (Java RMI); roles and rights management; feedback mechanisms
9.1.3	Development of an alternative, mobile iOS-based client.	Investigating alternative visualizations for arguments on small screens.	Public interfaces of LASAD (Webservice); flexible data format; possibility to create additional visualizations
9.2.1	Evaluating the diagnosticity of argument diagrams.	Evaluate the role of arguments modeled in argumentation environments as means for diagnosing argumentation.	Configuration mechanisms; collaboration; roles and rights management; export function; feedback support
9.2.2	Combining Facebook with LASAD.	Enhancement of discussions in social networks via argument graphs.	Configuration mechanisms; roles and rights management
9.2.3	Deepening argumentative discourse via scripted roles.	Evaluation of the effect of roles in combination with sentence openers in the chat on the argumentative discourse.	Configuration mechanisms; roles and rights management; extensibility of interface elements
9.2.4	Use LASAD as study tool.	Learning how to conduct a study as part of a university course.	Configuration mechanisms; authoring tool
9.2.5	Using LASAD as discussion part of a learning concept that involves multiple tools.	Creating a CSCL environment to effectively teach 12 to 16-years-old students science and mathematics.	Configuration mechanisms; authoring tool; extensibility of ontology elements; flexible data format; communication with external tools; feedback support
9.2.6	Integrating the peer review system SWoRD with LASAD.	Evaluate the effect of combining peer reviews with argument maps and intelligent feedback.	Configuration mechanisms; authoring tool; communication with external tools; export function; feedback support

Table 9.1: Summary of external studies with LASAD

Overall, the results confirmed the impressions resulting from the internal evaluation (see chapter 8) of the LASAD framework.

As part of these external applications, the LASAD framework has been used in research (→ sections 9.1.2, 9.1.3b, 9.2.1, 9.2.2 & 9.2.3), teaching (→ sections 9.2.4 & 9.2.5) and development (→ sections 9.1.1, 9.1.2, 9.1.3a, 9.1.3b, 9.2.3 & 9.2.5) contexts. These contexts, however, covered multiple domains including the law (→ sections 9.1.1, 9.1.3 & 9.2.1), psychology (→ section 9.2.2), ethics (→ section 9.2.3) as well as science and mathematics (→ section 9.2.5). Even though the results have not been fully evaluated while writing these lines, this highlights the domain-independence and flexibility of LASAD (cf. research question 3 in chapter 1.2) as well as its capability to deal with open research issues in computer-supported argumentation (cf. research question 4 in chapter 1.2). With respect to the former, in particular the application to unexpected scenarios such as logical puzzle solving show that the generic approach of LASAD is indeed beneficial.

On the development side, these applications provide practical evidences that the requirements collected in chapter 5 are fulfilled. This includes (a) the platform independence of the framework (the studies and projects ran on Windows, Linux, Mac OS and iOS), (b) the interoperability and integration of the framework with external systems (Metafora, SWoRD) as well as (c) the simple extensibility of framework (table-visualization, sentence opener chat for the ethics study, integration of complete web pages and applications in Metafora).

On the research and teaching side, these applications showed that even users who have not been directly involved in the development process of LASAD were able to actually use the system for their needs. This use includes the configuration of the system via the authoring tool as well as the actual discussion support in classroom via LASAD. Here, the authoring tool in connection with the browser based client, which did not require any installation, turned out to be an appropriate way to promote the actual use (cf. research question 5 in chapter 1.2).

However, the applications also showed how unpredictable (in a positive way) research is. Even though LASAD was designed to be used in various contexts and domains out-of-the-box, there are always requirements and functionalities that have not been considered and will require additional extensions in the future.

10 Conclusion, Summary & Outlook

The ability to argue is essential in many aspects of life. Even though this importance has been widely recognized by the research and teaching communities, the education of these skills is carelessly neglected. The reasons for it can be found, on the one hand, in the overall complexity of the ill-defined field of argumentation. Even though there are some guidelines to judge the quality of arguments, clear criteria what count as good argument do not exist. On the other hand, existing teaching methods to educate argumentation skills usually involve time-intensive face-to-face tutoring which is not applicable to larger groups or even whole classroom settings.

An approach that came up in the last years is the use of computers to support the education of argumentation skills. However, even though the variety of tools available is impressive, their effects are still largely unclear. While there is general agreement that these tools can be beneficial the implications and consequences of certain educative settings including factors such as argument representations, underlying argument model or even the role of collaboration are largely unexplored. This is a direct consequence of the variety of research fields involved in the field of computer-supported argumentation including, for example, researchers from the fields of educational psychology, artificial intelligence and computer science and the variety of domains in which argumentation takes place including, for example, business, science, medicine, the law or ethics to mention just a few.

Each of these groups is focusing on different aspects of computer-supported argumentation and systems that have been developed to investigate these aspects are oftentimes undocumented one-shot prototypes that cannot be reused in other contexts. Thus, research results are hardly comparable since the context in which they have been achieved differ in multiple factors.

In this thesis, the focus was on the computer science perspective on argumentation. The main contribution was the proposal and development of a highly flexible argumentation framework that can be used to (a) conduct valuable, more comparable research in the research field of computer-supported argumentation and (b) to educate argumentation skills. One of the key requirements of the framework was domain-independency. Therefore, this thesis highlighted that argumentation in different domains, even though differing in argumentation style and rules, share a common ground. This common ground has been explored in order to create a set of shared components that can be tailored to domain and context specific needs using a configuration approach and evaluated in concrete research and teaching contexts.

In more detail, the central question of this thesis was how a computer can provide domain-independent support for the education of argumentation skills. This question has been partitioned into five sub-questions.

The goal of the first sub-question was to identify similarities and differences of argumentation in various domains and to find out how these differences influence the design of computer-based tools that support argumentation. The answer provided in this thesis was split into a theoretical and a practical part.

The theoretical part started with a discussion of multiple definitions of argumentation (→ chapter 2.1). Based on these definition, a categorization into aggressive and productive argumentation was introduced. Whereas the former can be found in everyday life in situations in which inexperienced arguers try to make the best of it, the educational perspective is mostly absent. Thus, this thesis concentrated on the latter, that is, productive argumentation. Here, the educational perspective is twofold as described in chapter 2.2. On the one hand, there is the *learning to argue perspective*, which focuses on the education of skills that are essential to participate in reasonable argumentation. On the other hand, there is the *arguing to learn perspective*, which describes the application of argumentation as method to gain domain knowledge. However, both of them have to deal with the complex, ill-defined nature of argumentation which has been described in chapter 2.3. The main problem for teaching as well as for the concrete application of argumentation skills is the absence of a correct solution. Instead, multiple solutions are imaginable for most problems and the criteria that can be applied to judge the overall quality are not always clear. This, in turn, makes it unequally harder to teach argumentative knowledge compared to teaching in well-defined domains. However, in order to answer the first sub-question of this thesis, the theoretical background closes with a description of the characteristics of argumentation in the domains of science, law and ethics. This presentation highlighted domain-specific differences in argumentation and motivated the practical part of the answer to the research question.

The practical part dealt with the state-of-art in the area of computer-based argumentation. The system and literature review presented in chapter 3 revealed an impressive diversity of approaches to support argumentation and the education of argumentation skills developed within the last two decades. This diversity, however, is reflected in different characteristics including various kinds of argument representations (→ chapter 3.2) and underlying argumentation models (→ chapter 3.6), collaboration settings (→ chapter 3.4), interaction and process designs (→ chapters 3.3 & 3.5) as well as analysis and feedback techniques that benefit from artificial intelligence (→ chapter 3.7). It turned out, that these differences are not only present in systems and methods that were designed for different domains, but also in those which are explicitly designed to provide support for the same domains.

The second sub-question of this thesis was about the evaluation of practices that have shown to be effective in existing approaches to scaffold argumentation as well as those which failed and the reasons that caused it. Therefore, empirical studies that investigated the effects of these characteristics have been summarized (→ chapter

3.9). Even though the research done in this area indicates benefits of using computer support for argumentation and argumentation learning, the review made clear that the concrete effects of design decisions are still largely unclear. Thus, more — and, in particular, more specific — research is required in order to create an overarching theory how computers can be supportive in this area. As a first step, this thesis compiled a list of open research challenges in the area of computer-based argumentation. In order to master these challenges, it has been concluded that a generic and highly flexible argumentation framework is needed to allow for more systematic research in this area.

However, a literature and system review has a major limitation: It only captures the past. Current developments cannot be uncovered. Furthermore, the motivation behind design decisions is not always revealed. Therefore, the review has been extended by a survey among argumentation experts including researchers, teachers and developers from various domains such as business, ethics, science, law, medicine, math, etc (→ chapter 4). This survey indicated that some of the challenges extracted from the review have been recognized by the research community. Nevertheless, the survey confirmed the lack of research on effects of design decisions and, thus, confirmed the need for a generic argumentation framework that allows more systematic research.

The third sub-question of this thesis dealt with the question how divergent requirements and expectations of various domains and multiple groups can be met by a single framework. Based on the combined results of the review (→ chapter 3) and the survey (→ chapter 4), a list of requirements on a domain-independent generic argumentation framework which is capable to deal with the identified challenges has been compiled (→ chapter 5). However, the review as well as the survey revealed an almost total lack of system documentation and research publications about building generic, flexible, and reusable systems. Thus, an architecture proposal has been developed and presented in chapter 6 and a reference implementation was presented in chapter 7.

To evaluate the proposed solution, a combined approach consisting of internal (→ chapter 8) and external (→ chapter 9) evaluations was used. This way, the fourth sub-question, which asked whether the proposed framework is an appropriate means to conduct more comparable research of open issues in computer-supported argumentation, was answered positively.

The internal part involved two studies done by the author. The first one dealt with the, yet unclear, impact of ontology and collaboration on the outcomes of computer-based argumentation (→ chapter 8.1). By means of a controlled-lab study with 36 participants, a next step towards more comparable results has been done and the suitability of the framework's configuration approach has been shown. The most important results include that arguing in groups is indeed beneficial for the degree of elaboration of arguments and that simple ontologies are preferable compared to complex ones. The second study evaluated the usefulness of the authoring tool to enable inexperienced users to benefit from the framework's flexibility (→ chapter 8.2). Even though the results revealed some usability lacks, the overall result provided evidence that the authoring tool approach is able support inexperienced users in a reasonable way.

The external part of the evaluation reported applications of the proposed framework out of the developer's range. Thus, it highlighted the suitability of the approach to promote actual use in research and education. Here, assumed benefits of the framework has been confirmed. By means of additional clients (→ chapter 9.1.3), additional argument representations (→ chapters 9.1.2 & 9.1.3) and integrations with other systems (→ chapters 9.2.5 & 9.2.6) the flexibility and extensibility of the framework has been shown. In addition, successfully conducted studies in different contexts and domains highlighted the domain-independence of the framework (→ chapters 9.2.1, 9.2.2 & 9.2.3). Finally, the plans for future research by means of the proposed framework showed that it is capable to deal with open challenges in research.

With the fifth and final sub-question, which asked how the use of computer-based argumentation tools can be facilitated to promote their actual use in education, in mind a reference implementation of the architecture has been developed (→ chapter 7). By means of a web-based installation-free approach typical barriers caused by lacking technological usability has been avoided. In order to simplify the configuration of the system and, hence, to allow users who are not familiar with developments at all, an authoring tool has been introduced (→ chapter 7.4).

However, there are some limitations that should be kept in mind when evaluating the proposed framework and its underlying architecture. The challenge to provide domain-specific support implicates that extraordinary requirements that are only relevant to a single domain or a specific context cannot always be covered. That is, trade-offs are required. Especially in research contexts, new ideas are likely to generate additional needs. In addition, the proposed framework has neither been tested over a longer period of time nor with large amounts of users. Even though technical problems are unlikely, it might turn out that a long-term use of the system or the involvement of large user groups may reveal the need for additional support. However, the underlying architecture is designed to be extensible and, hence, prepared to be applicable to future challenges.

Nevertheless, this thesis made a significant step in the long process of understanding argumentation better in order to improve the education of highly important argumentation skills. However, there are three perspectives in which next steps are imaginable: (a) technology, (b) research and (c) teaching.

The first perspective is the technology perspective. Even though the technology and functionalities of the proposed LASAD framework have been carefully chosen based on the experiences from experts and existing systems, there is always something that was not considered as mentioned previously. Thus, the framework will be subject to change in the next years. Potential points that may be considered here include the technology used to implement the framework (e.g., HTML5 and NoSQL instead of JavaScript and a relational database), the functionalities of the framework (e.g., additional analysis and feedback methods, a sophisticated scripting engine or the integration of audio and video communication) as well as the integration with mobile devices which experienced a rapid rise in the last years. However, the underlying architecture of the framework (→ chapter 6) is prepared for these extensions or changes.

The second perspective is the research perspective. The proposed framework has proven to be an adequate means to conduct valuable and comparable research. Thus, the prerequisites are give. In the next step, the role of certain factors that are already present in argumentation systems should be carefully evaluated. Here, the review (→ chapter 3) and the survey among argumentation experts (→ chapter 4) revealed a set of open issues that can be used as starting points. Once the impacts of these factors are clear, the exploration of additional concepts that are frequently used in face-to-face argumentation, for instance gestures and facial expressions should be considered.

The third and final perspective is the teaching perspective. The real challenge here is to bring the results achieved in research into the classrooms. Here, the proposed framework made a first step, that is, to simplify the use to a degree which allows even inexperienced users to actually use the system. The usability, however, can only be one aspect. The true challenge here is to convince teachers that the systems are indeed helpful.

Bibliography

Allwood, J. (2002). Bodily communication dimensions of expression and content. In B. Granström, D. House, & I. Karlsson (Eds.), *Multimodality in language and speech systems* (pp. 7–26). Kluwer.

Andriessen, J. (2005). Arguing to learn. In *The Cambridge Handbook of the Learning Sciences* (pp. 443–459). Cambridge University Press.

Andriessen, J., Baker, M., & Suthers, D. D., Eds. (2003). *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments*. Computer-Support Collaborative Learning Series. Springer.

Aronson, E. (1978). *The Jigsaw Classroom*. Sage Publications Inc.

Aronson, E. & Patnos, S. (2011). *Cooperation in the Classroom: The Jigsaw Method*. Printer & Martin Ltd.

Asterhan, C. S. C. & Schwarz, B. B. (2011). E-moderation of synchronous discussions in educational settings: A nascent practice. *The Journal of the Learning Sciences*, 20, 395–442.

Baghaei, N., Mitrovic, A., & Irwin, W. (2007). Supporting collaborative learning and problem-solving in a constraint-based CSCL environment for UML class diagrams. *International Journal of Computer-Supported Collaborative Learning*, 2(2-3), 159–190.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6), 574–594.

Bangor, A., Kortum, P. T., & Miller, J. T. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *International Journal of Usability Studies*, 4(3), 114–123.

Belgiorno, F., Chiara, R. D., Manno, I., Overdijk, M., Scarano, V., & van Diggelen, W. (2008). Face to face cooperation with CoFFEE. In P. Dillenbourg & M. Specht (Eds.), *Proceedings of the 3rd European Conference on Technology Enhanced Learning (EC-TEL 2008)* (pp. 49–57). Springer.

Bell, P. (1997). Using argument representations to make thinking visible for individuals and groups. In R. Hall, N. Miyake, & N. Enyedy (Eds.), *Proceedings of the 2nd International Conference on Computer Support for Collaborative Learning (CSCL 1997)* (pp. 10–19).

Bell, P. & Linn, M. C. (2000). Scientific arguments as learning artifacts: designing for learning from the web with KIE. *International Journal of Science Education*, 22(8), 797–817.

Bex, F., van den Braak, S., Oostendorp, H., Prakken, H., Verheij, B., & Vreeswijk, G. A. W. (2007). Sense-making software for crime investigation: How to combine stories and arguments? *Law, Probability and Risk*, 6(1-4), 145–168.

Bouyias, Y. N., Demetriadis, S. N., & Tsoukalas, I. A. (2008). iArgue: A web-based argumentation system supporting collaboration scripts with adaptable fading. In *Proceedings of the 8th International Conference on Advanced Learning Technologies (ICALT 2008)* (pp. 477–479).: IEEE Computer Society Press.

Bozdag, E., Mesbah, A., & van Deursen, A. (2007). A comparison of push and pull techniques for AJAX. Delft University of Technology Technical Report Series - Report TUD-SERG-2007-016a.

Brandes, U., Eiglsperger, M., Herman, I., Himsolt, M., & Mashall, M. S. (2002). GraphML progress report - structural layer proposal. In P. Mutzel, M. Jünger, & S. Leipert (Eds.), *Graph Drawing*, volume 2265 of *Lecture Notes in Computer Science* (pp. 109–112). Springer.

Brenner, J. (2011). Die Nutzung des LASAD-Systems zur Unterstützung juristischer Argumentation. Diploma Thesis. Available online: http://hcis.in.tu-clausthal.de/pubs/2010/tuc/die_nutzung_des_lasad_systems_zur_unterstutzung_juristischer_argumentation.pdf (last visited 2011-07-15).

Brooke, J. (1996). SUS - a quick and dirty usability scale. In P. W. Jordan, B. Thomas, & B. A. Weerdmeester (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.

Buckingham Shum, S. J., MacLean, A., Bellotti, V. M. E., & Hammond, N. V. (1997). Graphical argumentation and design cognition. *Human-Computer Interaction*, 12(3), 267–300.

Buckingham Shum, S. J., Selvin, A. M., Sierhuis, M., Conklin, J., Haley, C. B., & Nuseibeh, B. (2006). Hypermedia support for argumentation-based rationale: 15 years on from gIBIS and QOC. In A. H. Dutoit, R. McCall, & I. Mistrik (Eds.), *Rationale Management in Software Engineering* (pp. 111–132). Springer.

Buckingham Shum, S. J., Uren, V., Li, G., Domingue, J., Motta, E., & Mancini, C. (2002). Designing representational coherence into an infrastructure for collective sense-making. Invited discussion paper presented at the 2nd International Workshop on Infrastructures for Distributed Collective Practices.

Carr, C. S. (2003). Using computer supported argument visualization to teach legal argumentation. In P. A. Kirschner, S. J. Buckingham Shum, & C. S. Carr (Eds.), *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making* (pp. 75–96). Springer.

Chang, F., Dean, J., Ghemawat, S., Hsieh, W. C., Wallach, D. A., Burrows, M., Chandra, T., Fikes, A., & Gruber, R. E. (2006). Bigtable: a distributed storage system for structured data. In *Proceedings of the 7th USENIX Symposium on Operating Systems Design and Implementation - Volume 7*, OSDI '06 Berkeley, CA, USA: USENIX Association.

Chesñevar, C., McGinnis, J., Modgil, S., Rahwan, I., Reed, C., Simari, G., South, M., Vreeswijk, G., & Willmott, S. (2006). Towards an argument interchange format. *The Knowledge Engineering Review*, 21(4), 293–316.

Cho, K. & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education*, 48(3), 409–426.

Cho, K.-L. & Jonassen, D. H. (2003). The effects of argumentation scaffolds on argumentation and problem solving. *Educational Technology: Research and Development*, 50(3), 5–22.

Chung, L. & do Prado Leite, J. C. S. (2009). On non-functional requirements in software engineering. In A. T. Borgida, V. K. Chaudhri, P. Giorgini, & E. Yu (Eds.), *Conceptual Modeling: Foundations and Applications*, volume 5600 of *Lecture Notes in Computer Science* (pp. 363–379). Springer.

Clark, H. H. & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (pp. 127–148). American Psychological Association.

Conklin, J. & Begeman, M. L. (1987). gIBIS: A hypertext tool for team design deliberation. In *Proceedings of the ACM Conference on Hypertext (1987)* (pp. 247–251).

Conklin, J. & Begeman, M. L. (1988). gIBIS: A hypertext tool for exploratory policy discussion. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work (1988)* (pp. 140–152).: ACM Press.

Conklin, J. & Begeman, M. L. (1989). gIBIS: A tool for all reasons. *Journal of the American Society for Information Science*, 40(3), 200–213.

Corbel, A., Girardot, J. J., & Jaillon, P. (2002). DREW: A dialogical reasoning web tool. In A. Méndez-Vilas & J. A. Mesa Gonzáles (Eds.), *Proceedings of the International Conference on Information and Communication Technologies in Education (ICTE2002)* (pp. 516–521).

de Groot, R., Drachman, R., Hever, R., Schwarz, B. B., Hoppe, U., Harrer, A., de Laat, M., Wegerif, R., McLaren, B. M., & Baurens, B. (2007). Computer supported moderation of e-discussions: the ARGUNAUT approach. In *Proceedings of the Conference on Computer-Supported Collaborative Learning (CSCL2007).*

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2010). Roles as a structuring tool in online discussion groups: The differential impact of different roles on social knowledge construction. *Computers in Human Behavior*, 26(4), 516–523.

DeCandia, G., Hastorun, D., Jampani, M., Kakulapati, G., Lakshman, A., Pilchin, A., Sivasubramanian, S., Vosshall, P., & Vogels, W. (2007). Dynamo: Amazon's highly available key-value store. In *Proceedings of 21st ACM SIGOPS symposium on Operating systems principles*, SOSP '07 (pp. 205–220). New York, NY, USA: ACM.

Devedzic, V. & Harrer, A. (2005). Software patterns in ITS architectures. *International Journal of Artificial Intelligence in Education*, 15(2), 63–94.

Dillenbourg, P. (2002). Overscripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL. Can we support CSCL?* (pp. 61–91). Open Universiteit Nederland.

Dillenbourg, P. & Hong, F. (2008). The mechanics of CSCL macro scripts. *International Journal of Computer-Supported Collaborative Learning (ijCSCL)*, 3(1), 5–23.

Dillenbourg, P. & Jermann, P. (2006). Designing integrative scripts. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), *Scripting Computer-Supported Collaborative Learning* (pp. 5–23). Springer.

Dragon, T., Gutierrez-Santos, S., Mavrikis, M., & McLaren, B. M. (2011a). Analysing and supporting students' interactions in synthesized learning environments: A case study with a microworld and a discussion tool. Workshop on Technology-Enhanced Learning for Mathematics and Science: Landmark Research and New Contributions (TELMAS@ECTEL2011).

Dragon, T., McLaren, B. M., Mavrikis, M., & Geraniou, E. (2011b). Scaffolding collaborative learning opportunities: Integrating microworld use and argumentation. International Workshop on Adaptive Support for Team Collaboration.

Easterday, M. W., Aleven, V., & Scheines, R. (2007). 'Tis better to construct than to receive? The effects of diagram tools on causal reasoning. In *Proceedings of the Conference on Artificial Intelligence in Education (AIED2007)* (pp. 93–100).

Einstein, A. (1954). *Ideas and opinions*. Crown Publishers, Inc.

Fischer, F., Kollar, I., Mandl, H., & Haake, J. M., Eds. (2006). *Scripting Computer-Supported Collaborative Learning*. Springer.

Fraternali, P., Rossi, G., & Sánchez-Figueroa, F. (2010). Rich internet applications. *IEEE Internet Computing*, (pp. 9–12).

Gamma, E., Helm, R., Johnson, R., & Vlissides, J., Eds. (1995). *Design Patterns. Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional Computing Series. Addison-Wesley.

Gehringer, E. F. (2001). Electronic peer review and peer grading in computer-science courses. In *Proceedings of the 32nd SIGCSE Technical Symposium on Computer Science Education* (pp. 139–143).

Glinz, M. (2007). On non-functional requirements. In *Proceedings of the 15th IEEE International Requirements Engineering Conference (RE2007)* (pp. 21–26).

Goodman, B. A., Linton, F. N., Gaimari, R. D., Hitzeman, J. M., Ross, H. J., & Zarrella, G. (2005a). Using dialogue features to predict trouble during collaborative learning. *User Modeling and User-Adapted Interaction*, 15(1-2), 85–134.

Goodman, B. A., Linton, F. N., Gaimari, R. D., Hitzeman, J. M., Ross, H. J., & Zarrella, G. (2005b). Using dialogue features to predict trouble during collaborative learning (Erratum). *User Modeling and User-Adapted Interaction*, 16(1), 83–84.

Gordon, T. F. (2007). Visualizing Carneades argument graphs. *Law, Probability and Risk*, 6(1-4), 109–117.

Gordon, T. F. & Karacapilidis, N. (1997). The Zeno argumentation framework. In *Proceedings of the 6th International Conference on Artificial Intelligence and LAW (ICAIL 1997)* (pp. 10–18).

Gruber, T. R. (1993). A translation approach to portable ontology specification. *Knowledge Acquisition*, 5(2), 199–220.

Hair, D. C. (1991). Legalese: A legal argumentation tool. *SIGCHI Bulletin*, 23(1), 71–74.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1).

Harrell, M. (2008). No computer program required: Even pencil-and-paper argument mapping improves critical thinking skills. *Teaching Philosophy*, 31(4), 351–374.

Harrer, A. & Devedzic, V. (2002). Design and analysis patterns in ITS architectures. In *Proceedings of the International Conference on Computers in Education (ICCE2002)*.

Herman, I. & Marshall, M. S. (2000). GraphXML - an xml based graph interchange format. Available online: http://www.cwi.nl/ftp/CWIreports/INS/INS-R0009.pdf, last visited: 2009-10-29.

Israel, J. & Aiken, R. (2007). Supporting collaborative learning with an intelligent web-based system. *International Journal of Artificial Intelligence in Education*, 17(1), 3–40.

Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education*, 38, 1217–1218.

Janssen, J., Erkens, G., Kirschner, P. A., & Kanselaar, G. (2010). Effects of representational guidance during computer-supported collaborative learning. *Instructional Science*, 38(1), 59–88.

Jeong, A. C. (2003). The sequential analysis of group interaction and critical thinking in online threaded discussions. *The American Journal of Distance Education*, 17(1), 25–43.

Jermann, P., Dillenbourg, P., & Brouze, J.-C. (1999). Dialectis for collective activities: an approach to virtual campus design. In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED1999).*

Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured learning outcomes. *Educational Technology: Research and Development*, 45(1), 65–94.

Jonassen, D. H. & Kim, B. (2010). Arguing to learn and learning to argue: design justifications and guidelines. *Educational Technology Research and Development*, 58(4), 439–457.

Karacapilidis, N. & Papadias, D. (2001). Computer supported argumentation and collaborative decision making: The Hermes system. *Information Systems*, 26, 259–277.

Karacapilidis, N., Tzagarakis, M., Karousos, N., Gkotsis, G., Kallistros, V., Christodoulou, S., Mettouris, C., & Nousia, D. (2009). Tackling cognitively-complex collaboration with CoPe_it! *International Journal of Web-Based Learning and Teaching Technologies*, 4(3), 22–38.

Keefer, M. W., Zeitz, C. M., & Resnick, L. B. (2000). Judging the quality of peer-led student dialogues. *Cognition and Instruction*, 18(1), 53–81.

Kerr, N. L. & Bruun, S. E. (1983). Dispensability of member effort and group motivation losses: Free-rider effects. *Journal of Personality and Social Psychology*, 44(1), 78–94.

Kim, J., Shaw, E., Ravi, S., Tavano, E., Arromratana, A., & Sarda, P. (2008). Scaffolding on-line discussions with past discussions: An analysis and pilot study of Pedabot. In B. P. Woolf, E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), *Proceedings of the 9th International Conference on Intelligent Tutoring Systems Conference (ITS2008)* (pp. 343–352).: Springer.

Klein, M. & Iandoli, L. (2008). Supporting collaborative deliberation using a large-scale argumentation system: The MIT collaboratorium. MIT Sloan Research Paper No. 4691-08. Available online: http://ssrn.com/abstract=1099082, last visited: 2010-08-02.

Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., & Fischer, F. (2007). Specifying computer-supported collaboration scripts. *Computer-Supported Collaborative Learning*, 2(2-3), 211–224.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts - a conceptual analysis. *Educational Psychology Review*, 18(2), 159–185.

Krug, S. (2006). *Don't make me think! Web Usability - Das intuitive Web*. mitp Business, 2nd rev. edition.

Krug, S. (2010). *Web Usability: Rocket Surgery Made Easy*. Addison-Wesley.

Kuhn, D. (1991). *The Skills of Argument*. Cambridge University Press.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, 77(3), 319–337.

Kumar, R., Rosé, C. P., Wang, Y.-C., Joshi, M., & Robinson, A. (2007). Tutorial dialogue as adaptive collaborative learning support. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), *Proceedings of the 13th International Conference on Artificial Intelligence in Education (AIED2007)* (pp. 383–390).: IOS.

Kunz, W. & Rittel, H. (1970). Issues as elements of information systems. Working paper #131. Institut für Grundlagen der Planung I.A. University of Stuttgart, Germany. Available online: http://www.cc.gatech.edu/~ellendo/rittel/rittel-issues.pdf, last visited: 2009-10-29.

Lakshman, A. & Malik, P. (2010). Cassandra: a decentralized structured storage system. *SIGOPS Oper. Syst. Rev.*, 44, 35–40.

Le, N.-T., Menzel, W., & Pinkwart, N. (2010). Considering ill-definedness of problem tasks under the aspect of solution space. In D. Wilson, C. Murray, & H. W. Guesgen (Eds.), *Proceedings of the 23rd International Conference of the Florida Artificial Intelligence Research Society (FLAIRS)* (pp. 534–535). Daytona Beach, FL, USA: AAAI.

Leavitt, N. (2010). Will NoSQL databases live up to their promise? *Computer*, 43(2), 12–14.

Leitão, S. (2000). The potential of argument in knowledge building. *Human Development*, 43(6), 332–360.

Linn, M. C., Clark, D., & Slotta, J. D. (2003). WISE design for knowledge integration. *Science Education*, 87(4), 517–538.

Loll, F. & Pinkwart, N. (2009a). CITUC: Automatisierte Lösungsbewertung im E-Learning durch kollaboratives Filtern. In H. R. Hansen, D. Karagiannis, & H. G. Fill (Eds.), *Tagungsband 2 der 9. Internationalen Tagung Wirtschaftsinformatik (WI)* (pp. 411–420). Vienna, Austria: Österreichische Computer Gesellschaft.

Loll, F. & Pinkwart, N. (2009b). Disburdening tutors in e-learning environments via web 2.0 techniques. In D. Dicheva, R. Mizoguchi, & J. Greer (Eds.), *Semantic Web Technologies for e-Learning* (pp. 279–298). IOS Press.

Loll, F. & Pinkwart, N. (2009c). Using collaborative filtering algorithms as elearning tools. In R. H. Sprague (Ed.), *Proceedings of the 42nd Hawaii International Conference on Ssytem Sciences (HICSS 2009).*

Loll, F. & Pinkwart, N. (2011). Guiding the process of argumentation: The effects of ontology and collaboration. In M. Spada, G. Stahl, N. Miyake, & N. Law (Eds.), *Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings. Volume I - Long Papers* (pp. 296–303).: International Society of the Learning Sciences.

Loll, F., Pinkwart, N., Scheuer, O., & McLaren, B. M. (2011a). Developing collaborative argumentation systems: What advice do the experts have? In M. Spada, G. Stahl, N. Miyake, & N. Law (Eds.), *Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings. Volume II - Short Papers & Posters* (pp. 980–981).: International Society of the Learning Sciences.

Loll, F., Pinkwart, N., Scheuer, O., & McLaren, B. M. (2011b). How tough should it be? Simplifying the development of argumentation systems using a configurable platform. In N. Pinkwart & B. M. McLaren (Eds.), *Educational Technologies for Teaching Argumentation Skills*. Bentham Science Publishers.

Loll, F., Scheuer, O., McLaren, B. M., & Pinkwart, N. (2010a). Computer-supported argumentation learning: A survey of teachers, researchers, and system developers. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Proceedings of the 5th European Conference on Technology Enhanced Learning*: Springer.

Loll, F., Scheuer, O., McLaren, B. M., & Pinkwart, N. (2010b). Learning to argue using computers - a view from teachers, researchers, and system developers. In V. Aleven, J. Kay, & J. Mostow (Eds.), *ITS 2010, Part II, LNCS 6095* (pp. 377–379).: Springer.

Lonchamp, J. (2010). CS in CSCL. In *Proceedings of the 13th International Conference on Interactive Computer Aided Learning (ICL2010).*

Loui, R. P., Norman, J., Altepeter, J., Pinkard, D., Linsday, J., & Foltz, M. (1997). Progress on Room 5: A testbed for public interactive semi-formal legal argumentation. In *Proceedings of the 6th International Conference on Artificial Intelligence and LAW (ICAIL 1997)* (pp. 207–214).

Lowrance, J., Harrison, I., Rodriguez, A., Yeh, E., Boyce, T., Murdock, J., Thomere, J., & Murray, K. (2008). Template-based structured argumentation. In A. Okada, S. Buckingham Shum, & T. Sherborne (Eds.), *Knowledge cartography: Software tools and mapping techniques* (pp. 307–333). Springer.

Lowrance, J. D. (2007). Graphical manipulation of evidence in structured arguments. *Law, Probability & Risk*, 6(1-4), 225–240.

Lund, K. (2007). The importance of gaze and gesture in interactive multimodal explanation. *Language Resources and Evaluation*, 41(3-4), 289–303.

Lund, K., Molinari, G., Séjourné, A., & Baker, M. (2007). How do argumentation diagrams compare when student pairs use them as a means for debate or as a tool for representing debate? *Computer-Supported Collaborative Learning*, 2(2-3), 273–295.

Lynch, C. (2011). PhD proposal: The diagnosticity of argument diagrams. Unpublished.

Lynch, C. (2012). *The Diagnosticity of Argument Diagrams*. PhD thesis, University of Pittsburgh.

Lynch, C., Ashley, K., Aleven, V., & Pinkwart, N. (2006). Defining "ill-defined domains"; a literature survey. In *Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains at the 8th International Conference on Intelligent Tutoring Systems (ITS)* (pp. 1–10).

Lynch, C., Ashley, K. D., Pinkwart, N., & Aleven, V. (2009). Concepts, structures, and goals: Redefining ill-definedness. *International Journal of Artificial Intelligence in Education*, 19(3), 253–266.

Malone, T. W. & Klein, M. (2007). Harnessing collective intelligence to address global climate change. *Innovations*, 2(3), 15–26.

Marshall, C. C., Halasz, F. G., Rogers, R. A., & Janssen, W. C. (1991). Aquanet: A hypertext tool to hold your knowledge in place. In *Proceedings of the 3rd annual ACM conference on Hypertext (HYPERTEXT '91)* (pp. 261–275).: ACM.

McAlister, S., Ravenscroft, A., & Scanlon, E. (2004). Combining interaction and context design to support collaborative argumentation using a tool for synchronous CMC. *Journal of Computer Assisted Learning*, 20, 194–204.

McLaren, B. M., Scheuer, O., & Mikšátko (2010). Supporting collaborative learning and e-discussions using artificial intelligence techniques. *International Journal of Artificial Intelligence in Education*, 20(1), 1–46.

McManus, M. M. & Aiken, R. M. (1995). Monitoring computer-based collaborative problem solving. *Journal of Artificial Intelligence in Education*, 6(4), 307–336.

Mitrovic, A. & Weerasinghe, A. (2009). Revisiting ill-definedness and the consequences for ITSs. In *Proceedings of the 14th International Conference on Artificial Intelligence in Education (AIED2009)* (pp. 375–382).: IOS Press.

Munneke, L., Andriessen, J., Kanselaar, G., & Kirschner, P. (2007). Supporting interactive argumentation: Influence of representational tools on discussing a wicked problem. *Computers in Human Behavior*, 23(3), 1072–1088.

Munneke, L., van Amelsvoort, M., & Andriessen, J. (2003). The role of diagrams in collaborative argumentation-based learning. *International Journal of Educational Research*, 39, 113–131.

Neu, B. (2010). Anpassung von argumentationsunterstützenden Systemen an die Anforderungen mobiler Endgeräte am Beispiel des iPod touch. Bachelor Thesis. Available online: http://hcis.in.tu-clausthal.de/pubs/2010/tuc/anpassung_von_argumentationsunterstutzenden_systemen_an_die_anforderungen_mobiler_endgerate_am_beispiel_des_ipod_touch.pdf (last visited 2011-07-15).

Newell, A. (1969). Heuristic programming: Ill-structured problems. *Progress in Operations Research*, 3, 361–413.

Newman, S. E. & Marshall, C. C. (1991). Pushing Toulmin too far: Learning from an argument representation scheme. Xerox Palo Alto Research Center Technical Report.

Nussbaum, E. M. & Edwards, O. V. (2011). Critical questions and argument stratagems: A framework for enhancing and analyzing students' reasoning practices. *The Journal of the Learning Sciences*, 20, 443–488.

Nussbaum, E. M., Winsor, D. L., Aqui, Y. M., & Poliquin, A. M. (2007). Putting the pieces together: Online argumentation vee diagrams enhance thinking during discussions. *International Journal of Computer-Supported Collaborative Learning (ijCSCL)*, 2(4), 479–500.

Okada, A. & Buckingham Shum, S. (2008). Evidence-based dialogue maps as a research tool to investigate the quality of school pupils' scientific argumentation. *International Journal of Research & Method in Education*, 31(3), 291–315.

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, 328(463), 463–466.

Pinkwart, N., Aleven, V., Ashley, K., & Lynch, C. (2006). Toward legal argument instruction with graph grammars and collaborative filtering techniques. In *Proceedings of the 8th International Conference on Intelligent Tutoring Systems (ITS2006)* (pp. 227–236).

Pinkwart, N., Aleven, V., Ashley, K., & Lynch, C. (2007). Evaluating legal argument instruction with graphical representations using LARGO. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), *Proceedings of the 13th International Conference on Artificial Intelligence in Education (AIED 2007)* (pp. 101–108).: IOS Press.

Pinkwart, N., Lynch, C., Ashley, K., & Aleven, V. (2008). Re-evaluating LARGO in the classroom: Are diagrams better than text for teaching argumentation skills? In B. P. Woolf, E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), *Proceedings of the 9th International Conference on Intelligent Tutoring Systems (ITS2008)* (pp. 90–100).: Springer.

Prakken, H. & Vreeswijk, G. (2002). Encoding schemes for a discourse support system for legal argument. 2nd Workshop in Computational Models of Natural Argument (ECAI2002), Lyon, France.

Rahwan, I. & Simari, G. R., Eds. (2007). *Argumentation in Artificial Intelligence*. Springer.

Ranney, M. & Schank, P. (1998). Toward an integration of the social and the scientific: Observing, modeling, and promoting the explanatory coherence of reasoning. In S. Read & L. Miller (Eds.), *Connectionist models of social reasoning and social behavior* (pp. 245–274). Erlbaum.

Ravenscroft, A., Sagar, M., Baur, E., & Oriogun, P. (2008). Ambient pedagogies, meaningful learning and social software. In S. Hatzipanagos & S. Warburton (Eds.), *Social software & developing community ontologies* (pp. 432–450). IGI Global.

Reed, C. & Rowe, G. (2004). Araucaria: Software for argument analysis, diagramming and representation. *International Journal of AI Tools*, 14(3-4), 961–980.

Reitman, W. R. (1964). Heuristic decision procedures, open constraints and the structure of ill-defined problems. In M. W. Shelly & G. L. Bryan (Eds.), *Human Judgments and Optimality* (pp. 282–315). John Wiley & Sons Inc.

Robertson, J., Good, J., & Pain, H. (1998). BetterBlether: The design and evaluation of a discussion tool for education. *International Journal of Artificial Intelligence in Education*, 9, 219–236.

Rolf, B. & Magnusson, C. (2002). Developing the art of argumentation. a software approach. In *Proceedings of the 5th International Conference on Argumentation (ISSA2002)* (pp. 919–926).

Roth, W.-M. (2000). From gesture to scientific language. *Journal of Pragmatics*, 32(11), 1683–1714.

Russell, A. (2006). Comet: Low latency data for the browser. Posting on blog: http://alex.dojotoolkit.org/?p=545.

Sampson, V. & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, 92(3), 447–472.

Sampson, V. & Clark, D. B. (2009). The impact of collaboration on the outcomes of scientific argumentation. *Science Education*, 93(3), 448–484.

Schank, P. & Ranney, M. (1995). Improved reasoning with Convince Me. In *CHI '95 Conference Companion* (pp. 276–277).: ACM, New York, NY.

Schank, P. K. (1995). *Computational Tools for Modeling and Aiding Reasoning: Assessing and Applying the Theory of Explanatory Coherence*. PhD thesis, University of California (Berkeley). Available online: http://dewey.soe.berkeley.edu/~schank/convinceme/papers/schank_dissy.ps.

Schellens, T., van Keer, H., de Wever, B., & Valcke, M. (2007). Scripting by assigning roles: Does it improve knowledge construction in asynchronous discussion groups? *International Journal of Computer-Supported Collaborative Learning (ijCSCL)*, 2(2-3), 225–246.

Schellens, T., van Keer, H., Valcke, M., & de Wever, B. (2005). Learning in asynchronous discussion groups: A multilevel approach to study the influence of student, group, and task characteristics. *Behaviour & Information Technology*, 36, 704–745.

Scheuer, O., Loll, F., McLaren, B. M., & Pinkwart, N. (2010). Computer-supported argumentation: A review of the state-of-the-art. *International Journal of Computer-Supported Collaborative Learning*, 5(1), 43–102.

Scheuer, O., McLaren, B. M., Loll, F., & Pinkwart, N. (2009). An analysis and feedback infrastructure for argumentation learning systems. In V. Dimitrova, R. Mizoguchi, B. du Bulay, & A. Graesser (Eds.), *Proceedings of the 14th International Conference on Artificial Intelligence in Education* (pp. 629–631).

Scheuer, O., McLaren, B. M., Loll, F., & Pinkwart, N. (2011). Automated analysis and feedback techniques to support argumentation: A survey. In N. Pinkwart & B. M. McLaren (Eds.), *Educational Technologies for Teaching Argumentation Skills*. Bentham Science Publishers.

Schneider, D. C., Voigt, C., & Betz, G. (2007). Argunet - a software tool for collaborative argumentation analysis and research. Workshop notes of the 7th Workshop on Computational Models of Natural Argument (CMNA VII); 20th International Joint Conference on Artificial Intelligence (IJCAI2007).

Schütt, T., Schintke, F., & Reinefeld, A. (2008). Scalaris: reliable transactional p2p key/value store. In *Proceedings of the 7th ACM SIGPLAN workshop on ERLANG*, ERLANG '08 (pp. 41–48). New York, NY, USA: ACM.

Schwarz, B., Drachman, R., & de Groot, R. (2010). Metafora — Learning to learn together: A visual language for social orchestration of educational activities. Project description. Available online: http://www.metafora-project.org/download/metafora_brochure.pdf (last visited 2011-03-14).

Schwarz, B. B. & Glassner, A. (2007). The role of floor control and of ontology in argumentative activities with discussion-based tools. *International Journal of Computer-Supported Collaborative Learning (ijCSCL)*, 2(4), 449–478.

Schwarz, B. B., Neuman, Y., & Biezunger, S. (2000). Two wrong may make a right... if they argue together! *Cognition and Instruction*, 18(4), 461–494.

Schworm, S. & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology*, 99(2), 285–296.

Shin, N., Jonassen, D. H., & McGee, S. (2003). Predictors of well-structured and ill-structured problem solving in an astronomy simulation. *Journal of Research in Science Teaching*, 40(1), 6–33.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189.

Simon, H. A. (1973). The structure of ill-structured problems. *Artificial Intelligence*, 4(3-4), 181–201.

Soller, A. (2001). Supporting social interaction in an intelligent collaborative learning system. *International Journal of Artificial Intelligence in Education*, 12, 40–62.

Soller, A. (2004). Computational modeling and analysis of knowledge sharing in collaborative distance learning. *User Modeling and User-Adapted Interaction*, 14(4), 351–381.

Soller, A., Monés, A. M., Jermann, P., & Mühlenbrock, M. (2005). From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. *International Journal on Artificial Intelligence in Education*, 15(4), 261–290.

Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1991). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. *Educational Technology*, 31(5), 24–33.

Stahl, G., Suthers, D. D., & Hesse, F. (2007). A double issue for CSCL 2007. *International Journal of Computer-Supported Collaborative Learning*, 2(2-3), 127–131.

Stegmann, K., Mu, J., Gehlen-Baum, V., & Fischer, F. (2011). The myth of over-scripting: Can novices be supported too much? In M. Spada, G. Stahl, N. Miyake, & N. Law (Eds.), *Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings. Volume I - Long Papers* (pp. 406–413).: International Society of the Learning Sciences.

Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning (ijCSCL)*, 2(4), 421–447.

Stonebraker, M. (2010). SQL databases v. NoSQL databases. *Communications of the ACM*, 53(4), 10–11.

Strijbos, J.-W., Martens, R. L., Jochems, W. M. G., & Broers, N. J. (2004). The effect of functional roles on group efficiency: Using multilevel modeling and content analysis to investigate computer-supported collaboration in small groups. *Small Group Research*, 35(2), 195–229.

Suthers, D. D. (2001). Architectures for computer supported collaborative learning. In *Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT2001)*.

Suthers, D. D. (2003). Representational guidance for collaborative inquiry. In J. Andriessen, M. Baker, & D. D. Suthers (Eds.), *Arguing to Learn*, Computer-Support Collaborative Learning Series (Vol. 1) (pp. 27–46). Springer.

Suthers, D. D., Connelly, J., Lesgold, A., Paolucci, M., Toth, E. E., Toth, J., & Weiner, A. (2001). Representational and advisory guidance for students learning scientific inquiry. In K. D. Forbus & P. J. Feltovich (Eds.), *Smart machines in education: The coming revolution in educational technology* (pp. 7–35). AAAI/MIT Press.

Suthers, D. D. & Hundhausen, C. D. (2003). An experimental study of the effects of representational guidance on collaborative learning processes. *The Journal Of The Learning Sciences*, 12(2), 183–218.

Suthers, D. D., Vatrapu, R., Medina, R., Joseph, S., & Dwyer, N. (2008). Beyond threaded discussion: Representational guidance in asynchronous collaborative learning environments. *Computers & Education*, 50(4), 1103–1127.

Suthers, D. D., Weiner, A., Connelly, J., & Paolucci, M. (1995). Belvedere: Engaging students in critical discussion of science and public policy issues. In *Proceedings of the 7th World Conference on Artificial Intelligence in Education (AIED 1995)* (pp. 266–273).

Taentzer, G. (2001). Towards common exchange formats for graphs and graph transformation systems. *Electronic Notes in Theoretical Computer Science*, 44(4), 28–41.

Tedesco, P. (2003). MArCo: Building an artificial conflict mediator to support group planning interactions. *International Journal of Artificial Intelligence in Education*, 13, 117–155.

Toth, E. E., Suthers, D. D., & Lesgold, A. M. (2001). “Mapping to Know”: The effects of representational guidance and reflective assessment on scientific inquiry. *Science Education*, 86(2), 264–286.

Toulmin, S. E. (2003). *The Uses of Argument.* Cambridge University Press, 2nd rev. edition.

Tsovaltzi, D., Rummel, N., McLaren, B. M., Pinkwart, N., Scheuer, O., Harrer, A., & Braun, I. (2010). Extending a virtual chemistry laboratory with a collaboration script to promote conceptual learning. *International Journal of Technology Enhanced Learning*, 2(1-2), 91–110.

van den Braak, S. W. & Vreeswijk, G. A. W. (2006). AVER: Argument visualization for evidential reasoning. In *Proceedings of the 9th Conference on Legal Knowledge and Information Systems (JURIX2006)* (pp. 151–156).

van Gelder, T. (2002). Argument mapping with Reason!Able. *APA Newsletter*, 2(1), 85–90.

van Gelder, T. (2003). Enhancing deliberation through computer supported visualization. In P. A. Kirschner, S. J. Buckingham Shum, & C. S. Carr (Eds.), *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making* (pp. 97–115). Springer.

van Gelder, T. (2007). The rationale for Rationale™. *Law, Probability and Risk*, 6, 23–42.

Verheij, B. (1998). Argue! - an implemented system for computer-mediated defeasible argumentation. In H. L. Poutré & J. van den Herik (Eds.), *Proceedings of the 10th Netherland/Belgium Conference on Artificial Intelligence (NAIC 1998)* (pp. 57–66).: CWI, Amsterdam.

Verheij, B. (2003). Artificial argument assistants for defeasible argumentation. *Artificial Intelligence*, 150, 291–324.

Vicknair, C., Macias, M., Zhao, Z., Nan, X., Chen, Y., & Wilkins, D. (2010). A comparison of a graph database and a relational database: a data provenance perspective. In *Proceedings of the 48th Annual Southeast Regional Conference*, ACM SE '10 New York, NY, USA: ACM.

Vizcaino, A., Contreras, J., Favela, J., & Prieto, M. (2000). An adaptive, collaborative environment to develop good habits in programming. In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Proceedings of the 5th International Conference on Intelligent Tutoring Systems (ITS 2000)* (pp. 262–271).: Springer.

von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching*, 45(1), 101–131.

von der Lieth Gardner, A. (1987). *An Artificial Intelligence Approach to Legal Reasoning*. MIT Press.

Voss, J. F. (2006). Toulmin's model and the solving of ill-structured problems. In D. Hitchcock & B. Verheij (Eds.), *Arguing on the Toulmin Model: New Essays in Argument Analysis and Evaluation* (pp. 303–311). Springer.

Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. (1983). Problem solving skill in the social sciences. *The Psychology of Learning and Motivation*, 17, 165–215.

Walton, D. N., Reed, C., & Macagno, F. (2008). *Argumentation Schemes*. Cambridge University Press.

Webb, N. M. & Palincsar, A. S. (1996). Group processes in the classroom. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 841–873). Simon & Schuster Macmillan.

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science*, 33(1), 1–30.

Wenger, E. (1987). *Artificial intelligence and tutoring systems.* Kaufmann.

Wigmore, J. H. (1931). *The principles of judicial proof.* Little, Brown & Co., 2nd edition.

Woolf, B. P., Murray, T., Marshall, D., Dragon, T., Kohler, K., Mattingly, M., Bruno, M., Murray, D., & Sammons, J. (2005). Critical thinking environments for science education. In *Proceedings of the 12th International Conference on Artificial Intelligence and Education (AIED 2005)* (pp. 702–709).

Zeidler, D. L. (1997). The central role of fallacious thinking in science education. *Science Education*, 81, 483–496.

Zhao, E. (2011). Tabellarische Visualisierung von Argumenten im LASAD-System. Unpublished Diploma Thesis.

Appendix

A List of Argumentation Systems, Methods and Studies

The review in chapter 3 covered the systems, methods, and studies shown in the table below. In the rightmost column, in brackets, the number of citations to the main paper of each system, method or study is provided, based on a Google Scholar[1] as an indicator of the influence of each system. This Google search was done at August 9th, 2011. All URLs were last visited on the same date.

No.	Tool	Feature description	Reference [#]
1	Academic Talk	collaborative, educational, sentence openers, based on dialogue game theory	McAlister et al. (2004) [89], http://www.londonmet.ac.uk/ltri/research/projects/at.htm
2	Aquanet	collaborative, configurable ontology	Marshall et al. (1991) [229]
3	Araucaria	transcript, argument schemes, central database for argument exchange	Reed & Rowe (2004) [156], http://araucaria.computing.dundee.ac.uk
4	Argue/ArguMed	argument assistance, legal domain	Verheij (2003) [92], http://www.ai.rug.nl/~verheij/aaa/argumed3.htm
5	ArguNet	collaborative, web-based	Schneider et al. (2007) [2], http://www.argunet.org
6	ARGUNAUT	educational, support system for human moderators, used with Digalo	de Groot et al. (2007) [24] McLaren et al. (2010) [13], http://www.argunaut.org
7	Athena	educational, report generator	Rolf & Magnusson (2002) [35], http://www.athenasoft.org
8	AVER	criminal investigations	van den Braak & Vreeswijk (2006) [8]
9	AVERs	criminal investigations	Bex et al. (2007) [26]
10	Belvedere v1&v2	educational, collaborative, ITS, scientific / evidential reasoning	Suthers et al. (1995) [178] Suthers et al. (2001) [56]
11	Belvedere v3&v4	educational, collaborative, multiple views, scientific / evidential reasoning	Suthers (2003) [59], http://lilt.ics.hawaii.edu/lilt/software/belvedere
12	BetterBlether	educational, collaborative, sentence openers	Robertson et al. (1998) [78]

[1] http://scholar.google.com

No.	Tool	Feature description	Reference [#]
13	Carneades	support of multiple proof-standards, IBIS	Gordon (2007) [19], http://carneades.berlios.de
14	CoChemEx	educational, collaborative, inquiry learning, chemistry, scripted	Tsovaltzi et al. (2010) [9]
15	CoFFEE	educational, collaborative, multiple tools, configurable	Belgiorno et al. (2008) [7], http://www.coffee-soft.org
16	Collaboratium	collaborative, IBIS	Klein & Iandoli (2008) [4], Malone & Klein (2007) [25], http://cci.mit.edu/research/climate.html
17	Collect-UML	educational, collaborative, problem solving, UML diagrams, ITS	Baghaei et al. (2007) [29]
18	Compendium	successor of Questmap, collaborative, IBIS	Buckingham Shum et al. (2006) [78], Okada & Buckingham Shum (2008) [5], http://compendium.open.ac.uk
19	Convince Me	educational, model of coherent reasoning	Ranney & Schank (1998) [34], Schank (1995) [18], http://www.soe.berkeley.edu/~schank/convinceme
20	CoPe_it!	successor of Hermes, (also) educational, collaborative, multiple views, support of multiple proof-standards, decision support, IBIS	Karacapilidis et al. (2009) [5]
21	CycleTalk Chat Environment	educational, collaborative, problem solving, thermodynamics, tutorial dialogues	Kumar et al. (2007) [70]
22	DebateGraph	collaborative, local views	http://www.debategraph.org
23	Debatepedia	collaborative, wiki-based	http://wiki.idebate.org
24	Digalo	educational, collaborative, configurable ontology	Schwarz & Glassner (2007) [21], http://www.dunes.gr
25	DREW	educational, collaborative, multiple tools	Corbel et al. (2002) [29]
26	Epsilon (with tutorial agent Pierce)	educational, collaborative, problem solving, OMT diagrams, sentence openers, interaction analysis, tutorial feedback, group and student model	Goodman et al. (2005a,b) [41]
27	Epsilon (interaction analysis)	educational, collaborative, problem solving, OMT diagrams, sentence openers, interaction analysis	Soller (2001) [270], Soller (2004) [61]
28	Group Leader Tutor	educational, collaborative, sentence openers, Group Leader agent to facilitate interaction	McManus & Aiken (1995) [138], Israel & Aiken (2007) [19]
29	Hermes	collaborative, support of multiple proof-standards, decision support, IBIS	Karacapilidis & Papadias (2001) [183], http://www-sop.inria.fr/aid/hermes
30	IBIS / gIBIS	collaborative, notational support to solve wicked problems	Conklin & Begeman (1988) [1426]
31	iLogos	educational, causal diagrams	Easterday et al. (2007) [10], http://www.phil.cmu.edu/projects/argument_mapping

No.	Tool	Feature description	Reference [#]
32	iArgue	web-based, scripts	Bouyias et al. (2008) [4]
33	Interloc	successor of Academic Talk, educational, collaborative, sentence openers, configurable dialogue games	Ravenscroft et al. (2008) [6], http://www.interloc.org
34	KIE / Sense-Maker, WISE	educational, container visualization, inquiry learning, science learning	Bell (1997) [148], Bell & Linn (2000) [297], Linn et al. (2003) [137], http://tels.sourceforge.net/sensemaker
35	LARGO	educational, legal argumentation, transcript, ITS	Pinkwart et al. (2006) [32], Pinkwart et al. (2007) [24], Pinkwart et al. (2008) [13]
36	Legalese	legal argumentation	Hair (1991) [11]
37	Pedabot	educational, support for technical discussion boards by IR	Kim et al. (2008) [9]
38	ProSupport	legal argumentation, form-based	Prakken & Vreeswijk (2002) [8]
39	Questmap	collaborative, IBIS	Carr (2003) [44]
40	Rashi/Human Biology Inquiry Tutor	educational, ITS, inquiry learning, multiple tools	Woolf et al. (2005) [8]
41	Rationale	educational, multiple argument modes	van Gelder (2007) [31], http://rationale.austhink.com
42	Reason!Able	educational	van Gelder (2002) [30], van Gelder (2003) [58]
43	Room 5	collaborative, legal argumentation, implements dialogue game	Loui et al. (1997) [53]
44	SEAS	decision support, argument templates, table, starburst and constellation depictions of multidimensional arguments	Lowrance (2007) [5], Lowrance et al. (2008) [6], http://www.ai.sri.com/~seas/
45	TC3	educational, collaborative, tool suite to support collaborative writing of argumentative texts	Munneke et al. (2003) [20]
46	Theorymaps.com	scientific argumentation, web-based	http://theorymaps.com/
47	VCRI	collaborative, educational	http://edugate.fss.uu.nl/~crocicl/vcri_eng.html
48	Zeno	predecessor of Carneades and Hermes, support of multiple proof-standards, decision support, IBIS	Gordon & Karacapilidis (1997) [191]
49	-	educational, collaborative	Jeong (2003) [175]
50	-	educational, collaborative, argument vee diagrams	Nussbaum et al. (2007) [13]
51	-	educational, collaborative, scripting by roles approach	Schellens et al. (2007) [19]
52	-	educational, collaborative, micro-scripting, Toulmin-based	Stegmann et al. (2007) [46]
53	-	educational, collaborative, integration of conceptual and discourse representations, uses Belvedere	Suthers et al. (2008) [78]
54	-	educational, collaborative, scientific argumentation	Sampson & Clark (2009) [36]

No.	Tool	Feature description	Reference [#]
55	-	educational, argument analysis, argument mapping	Harrell (2008) [3]
56	-	educational, representational guidance, argument representations, scientific argumentation	Suthers & Hundhausen (2003) [187]
57	-	educational, collaborative, argument diagramming	Lund et al. (2007) [22]

Table A.1: Overview of reviewed tools, methods and evaluations

B Argumentation Questionnaire

B.1 Personal Background Questions

1. I have experience with *researching* how people argue in my domain of expertise.

   Scale: 1 = Very little, 2 = Little, 3 = Somewhat, 4 = A lot, 5 = I am an expert

2. I have experience with *teaching* students to argue.

   Scale: 1 = Very little, 2 = Little, 3 = Somewhat, 4 = A lot, 5 = I am an expert

3. I have experience with *designing and/or developing software* to teach or support students in arguing.

   Scale: 1 = Very little, 2 = Little, 3 = Somewhat, 4 = A lot, 5 = I am an expert

4. I have used computer tools that support argumentation.

   Scale: 1 = Very little, 2 = Little, 3 = Somewhat, 4 = A lot, 5 = I am an expert

5. I have knowledge of artificial intelligence technology.

   Scale: 1 = Very little, 2 = Little, 3 = Somewhat, 4 = A lot, 5 = I am an expert

6. Please select the domain in which you have the most experience with argumentation

   Options: The law, Ethics, Science, Business, ... (Other)

B.2 Multiple Choice Questions

To test for answer consistency, there were typically two items designed to measure the same thing, e.g. Q3 and Q11.

1. Which answer best describes the type of argumentation that is taught and/or used in your primary domain of interest?

   Scale: 1 = Informal (e.g., everyday conversation), 2 = Somewhat informal, 3 = Partly informal / partly formal, 4 = Somewhat formal, 5 = Formal (e.g., formal proofs), 6 = I do not know

2. In my primary domain of interest, it is important that people learn argumentation through discussions, rather than on their own (e.g., from a book, by sketching arguments on paper, etc.).

   Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree, 6 = I do not know

3. I think that an argument represented in visual fashion – for instance, as a graph of participants' points and their relationships – can be very helpful to people in understanding and reflecting upon the argument.

   Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree, 6 = I do not know

4. The kind of argumentation I am interested in can be facilitated or supported by a computer system well enough so that the participants do not need to speak and argue face-to-face.

   Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree, 6 = I do not know

5. To what extent do rules of inference and/or axioms provide structure to arguments in your primary domain of interest?

   Scale: 1 = Very little, 2 = A little, 3 = Somewhat, 4 = A lot, 5 = A very great deal, 6 = I do not know

6. The rules and valid forms of argumentation can be learned in my domain of interest through individual study and practice.

   Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree, 6 = I do not know

7. In my primary domain weaknesses and errors in arguments can be identified by general and recurring patterns.

   Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree, 6 = I do not know

8. In my domain of interest it would be possible to define a set of rules that could be used to automatically detect general and recurring patters in arguments.

   Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree, 6 = I do not know

9. It is possible to asses the quality of an argument just by analyzing general patterns in argument.

   Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree, 6 = I do not know

10. Feedback on errors and problems when engaging in argumentation learning is most effectively provided...

    Scale: 1 = Immediately after the error or problem occurs, 2 = After the argument is over, 3 = Only when the participants explicitly ask for feedback, 4 = ... (Other), 6 = I do not know

11. Arguments shown in graphical fashion are likely to be helpful in learning or understanding argumentation.

    Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree, 6 = I do not know

12. Computer systems have the potential to support people in conducting useful, valid arguments over the Internet, perhaps even improving upon standard, face-to-face discussion.

    Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree, 6 = I do not know

B.3 Open-ended Questions

Based on the respondents' self reported expertise, they were distributed to up to two of the following question categories. If they had expertise in all three categories, the categories with the highest expertise were chosen. If there was a split between categories, the participant was assigned randomly between these categories.

Category 1: Research

a. In your primary domain of interest, what criteria do you apply to decide whether an argument is good?

b. Imagine that you have a software tool with graphical components representing different plausible argument moves that users can choose from. They might be able to choose from components such as claim, fact, or rebuttal and then fill in the selected shapes with text specific to their idea. Do you think such an approach would help or hinder users as they construct arguments and why?

c. Can you describe the typical process and roles that are used in arguments (or debates between parties) in your primary domain of interest?

Category 2: Teaching

a. In your primary domain of interest, what are the most common mistakes made by students (or typical misconceptions) in formulating arguments on their own (i.e., individually)?

b. What kinds of problems occur most frequently when your students practice argumentation collaboratively with one another? (As opposed to composing arguments on their own, as per question 1 above.)

c. What general types of feedback do you (or would you) give to students when they make oral or written arguments in your primary domain of interest?

Category 3: Design & Development

a. In the argumentation system (or systems) that you have designed and/or developed, what would you say was the best feature? Why?

b. Can you briefly describe the types of flexibility and/or configuration you provided in the design of your argumentation system (or systems)?

c. What would you say was the most important lesson (or lessons) learned in designing, developing and/or testing your argumentation system?

C Taxonomies

In this section the taxonomies that have been created during the coding process described in section 4.2.1 are presented. Single answers may be contained in multiple categories. The number n indicating the overall amounts of answers in the respective category does include all answers of the sub-categories. The background color (ranging from red to green) indicate the percentage of answers of the upper level category that are incorporated into the sub-category. In some cases the number of empty responses is reported. These numbers have not been added to the overall number of answers.

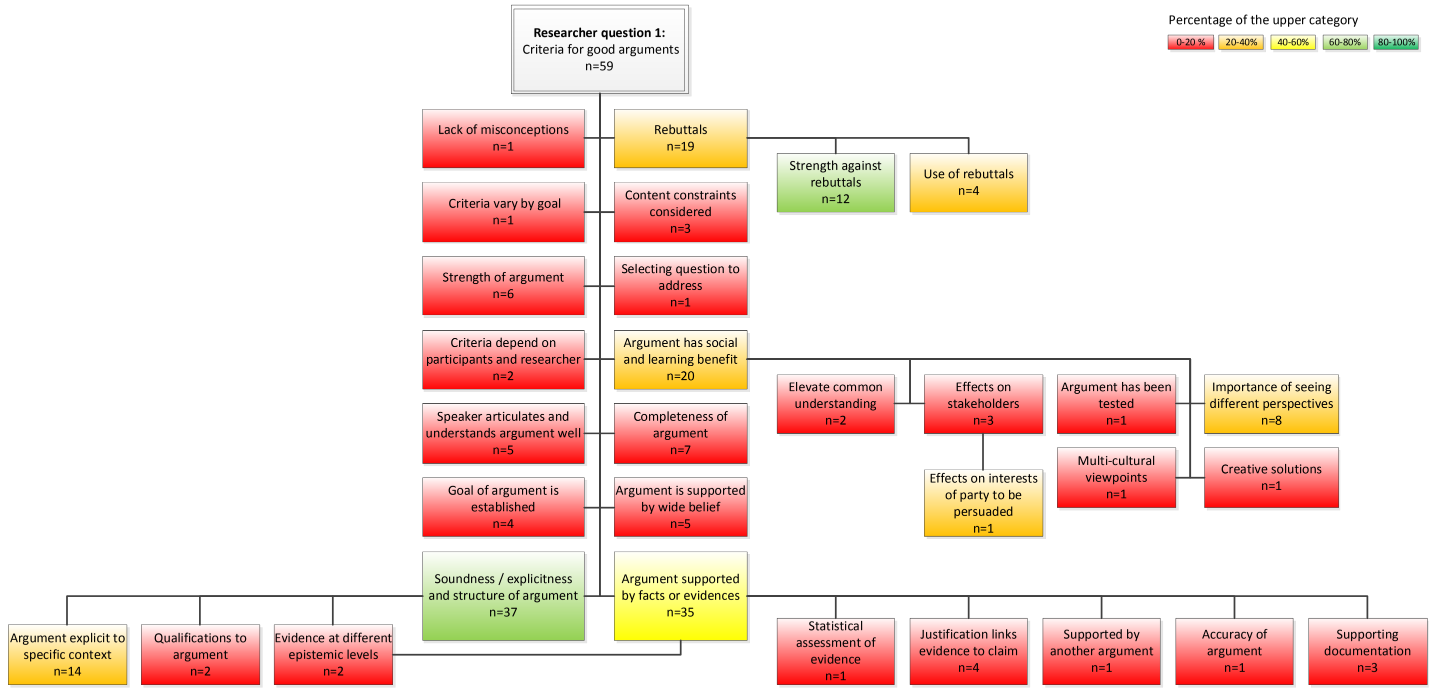


Figure C.1: Taxonomy of researcher question 1: In your primary domain of interest, what criteria do you apply to decide whether an argument is good?

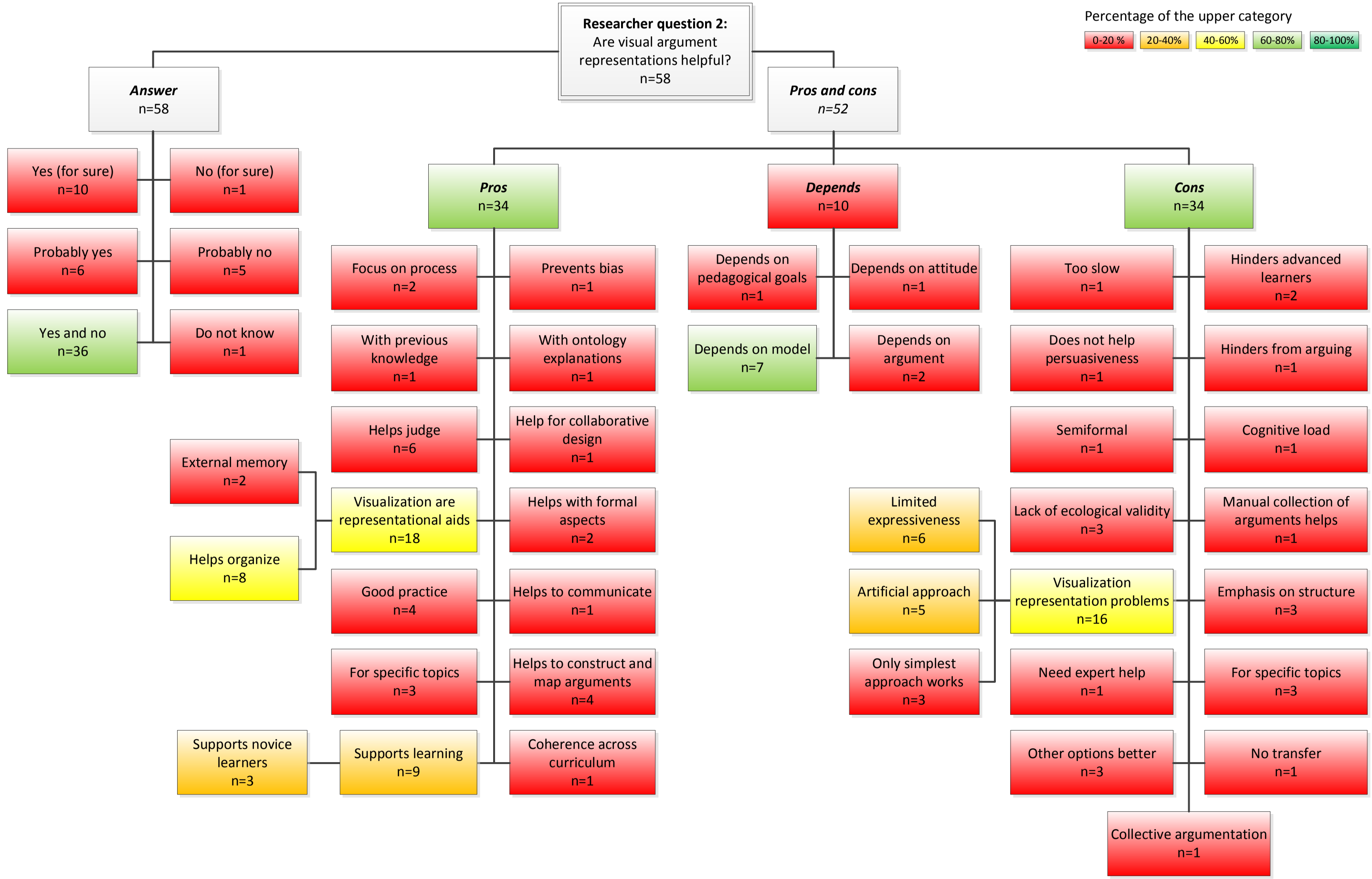


Figure C.2: Taxonomy of researcher question 2: Imagine that you have a software tool with graphical components representing different plausible argument moves that users can choose from. They might be able to choose from components such as claim, fact, or rebuttal and then fill in the selected shapes with text specific to their idea. Do you think such an approach would help or hinder users as they construct arguments and why?

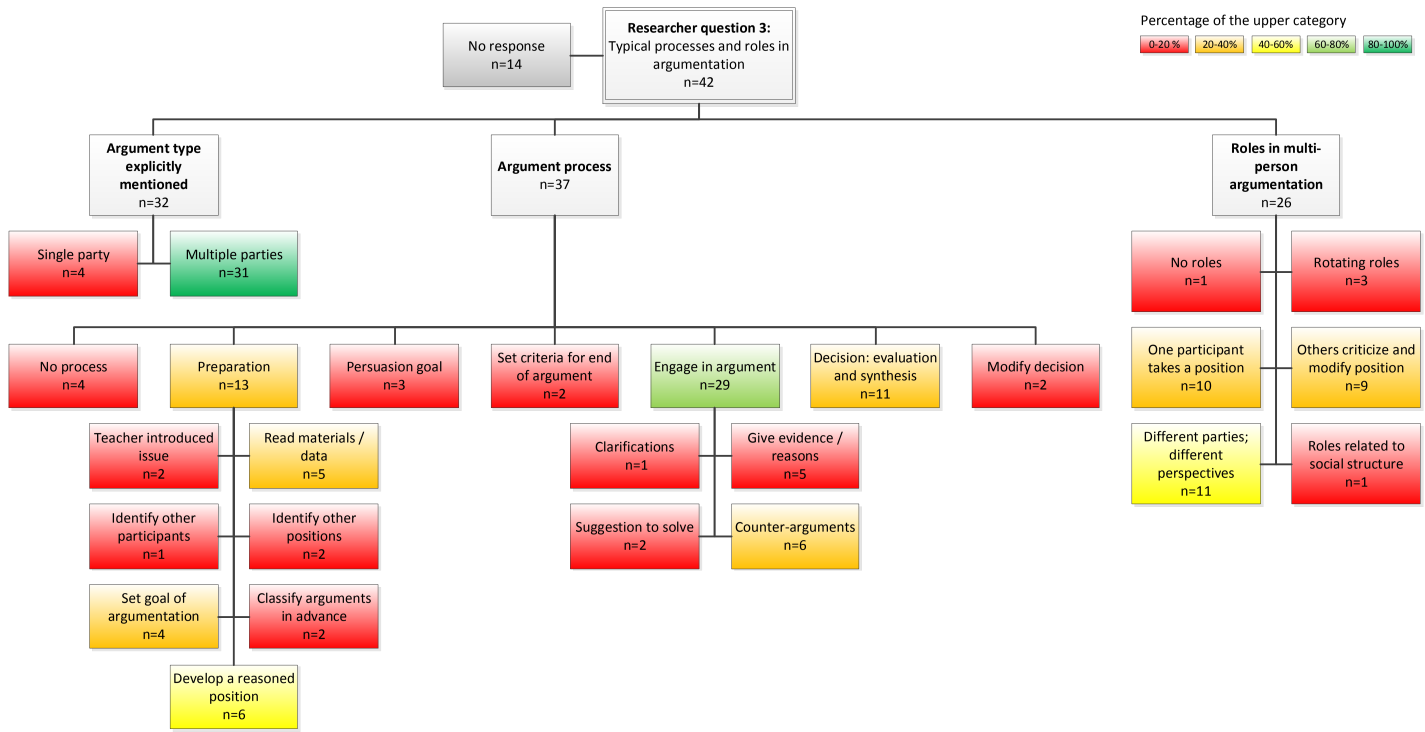


Figure C.3: Taxonomy of researcher question 3: Can you describe the typical process and roles that are used in arguments (or debates between parties) in your primary domain of interest?

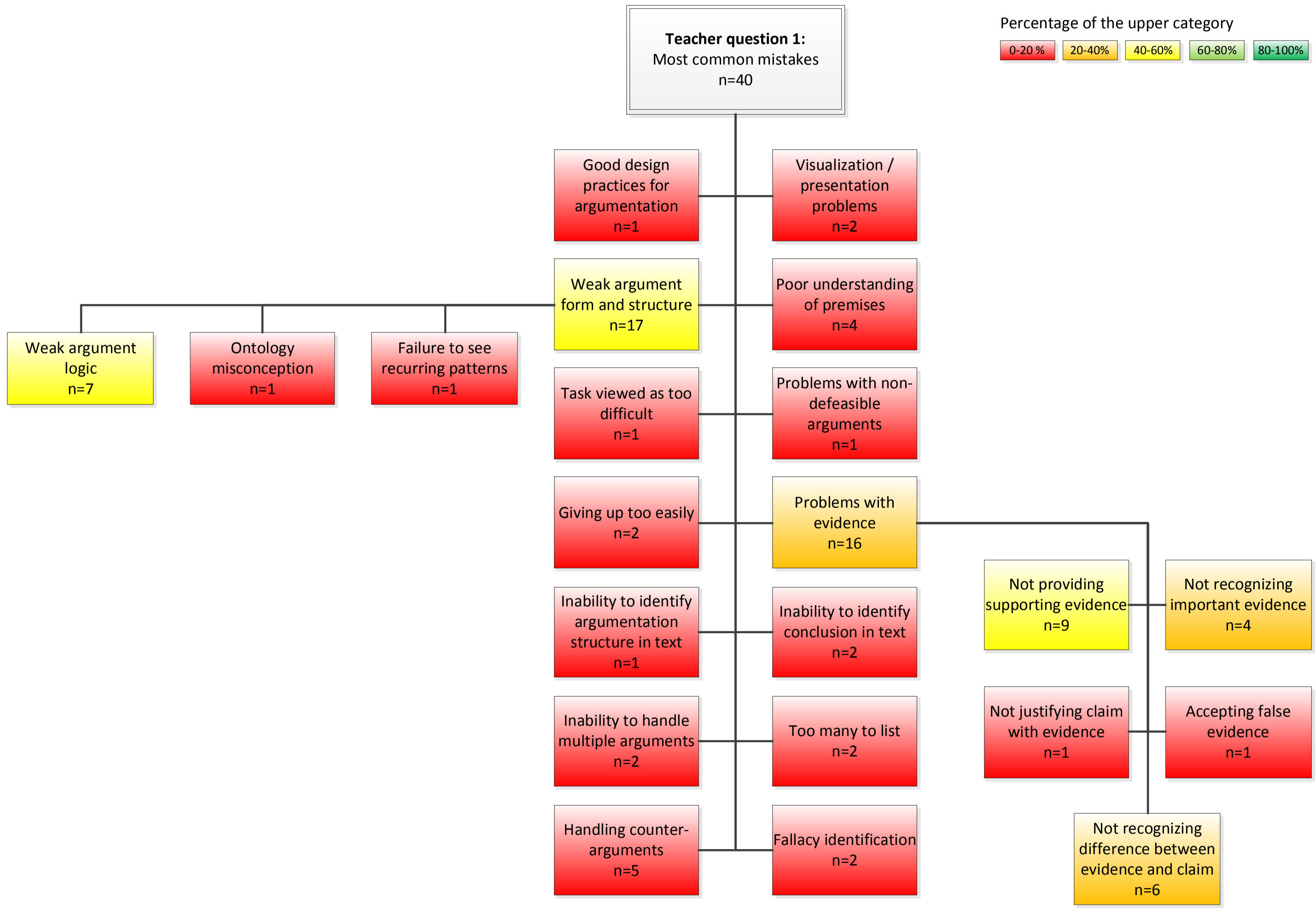


Figure C.4: Taxonomy of teacher question 1: In your primary domain of interest, what are the most common mistakes made by students (or typical misconceptions) in formulating arguments on their own (i.e., individually)?

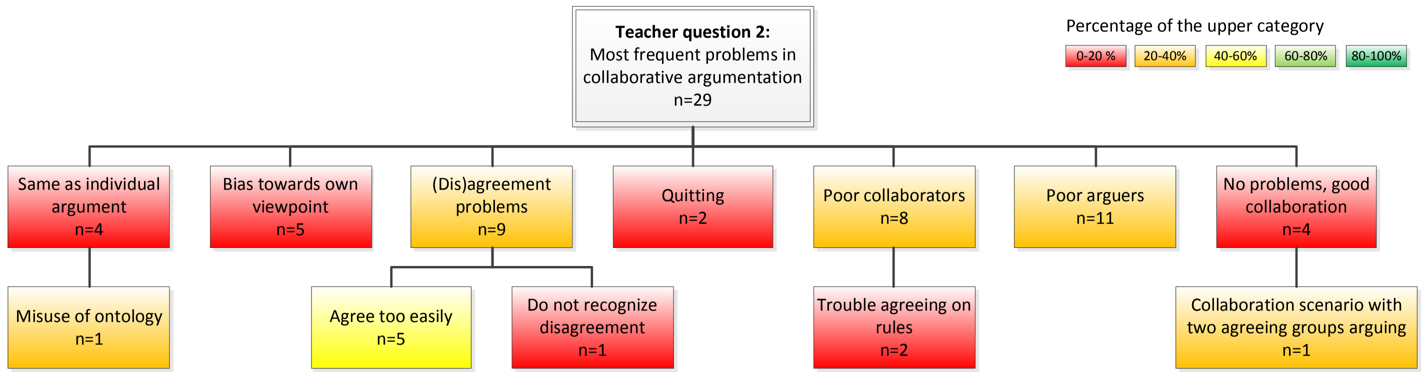


Figure C.5: Taxonomy of teacher question 2: What kinds of problems occur most frequently when your students practice argumentation collaboratively with one another?

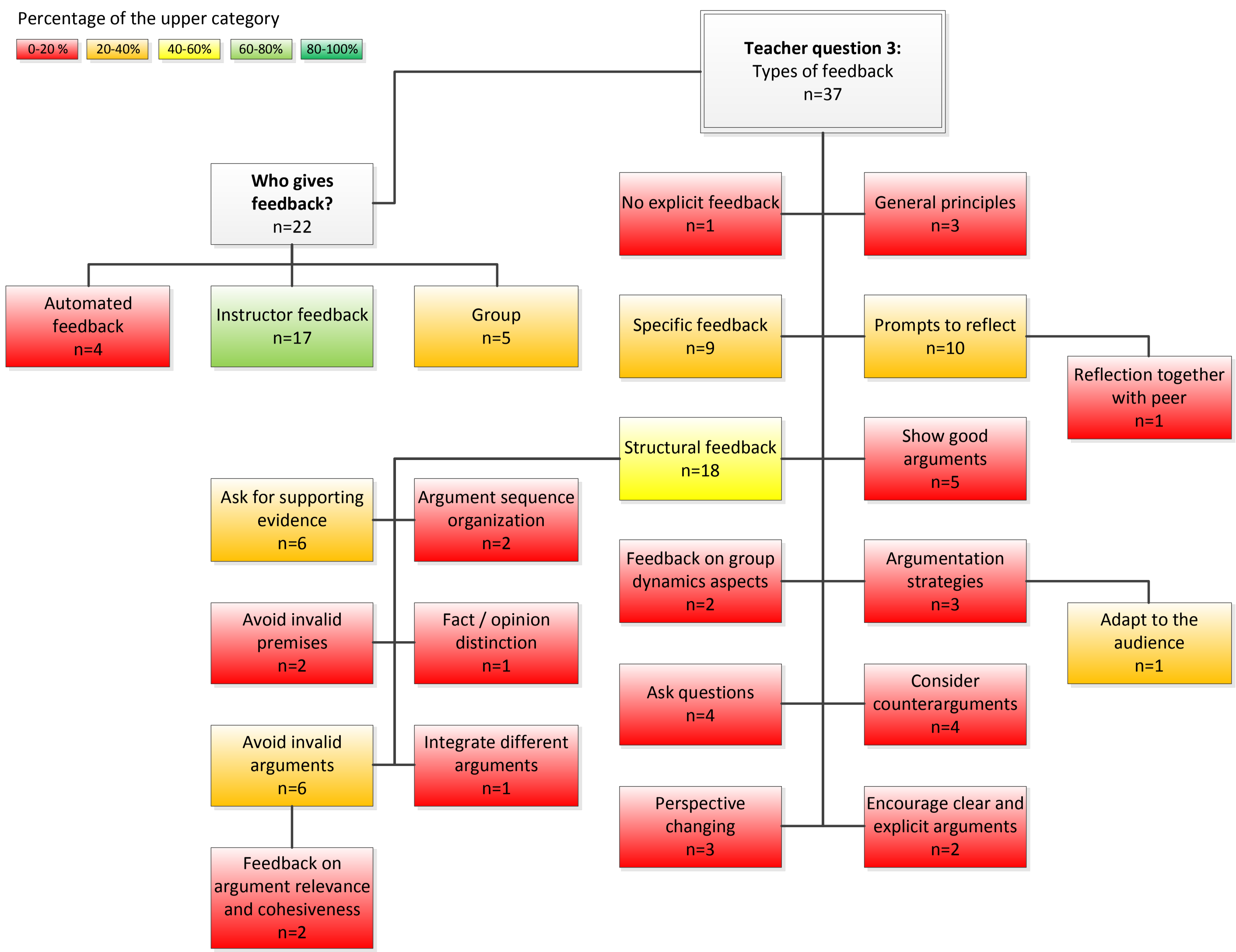


Figure C.6: Taxonomy of teacher question 3: What general types of feedback do you (or would you) give to students when they make oral or written arguments in your primary domain of interest?

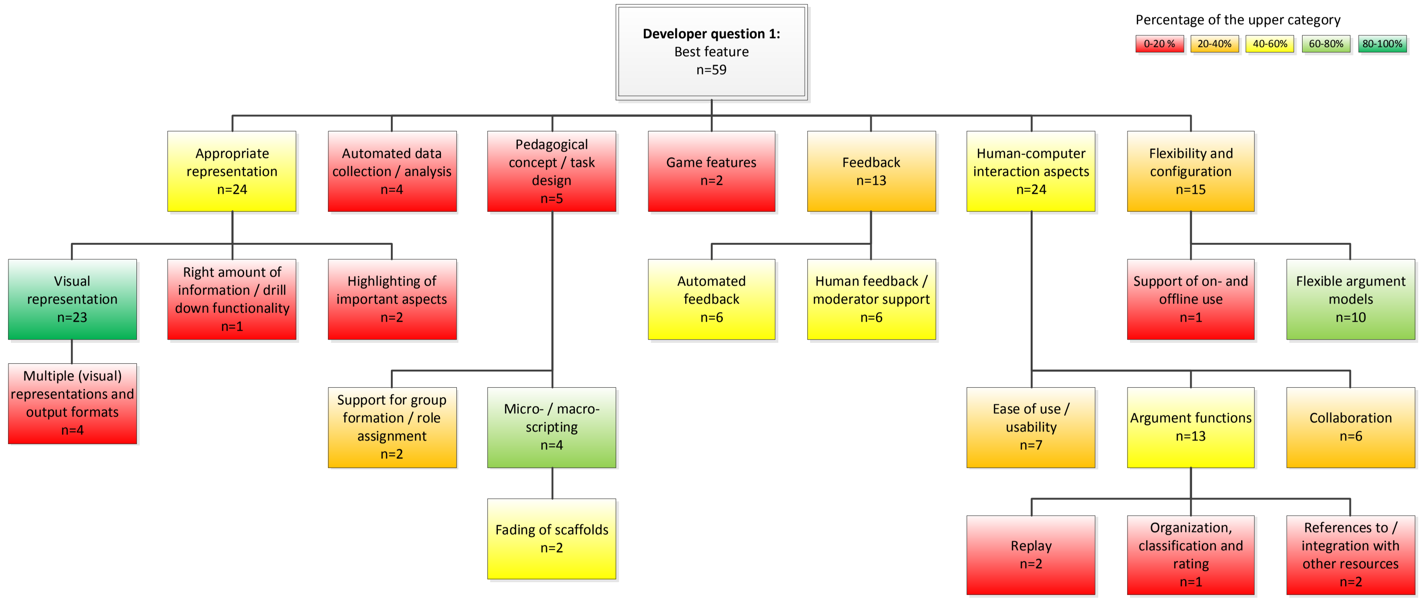


Figure C.7: Taxonomy of developer question 1: In the argumentation system (or systems) that you have designed and/or developed, what would you say was the best feature? Why?

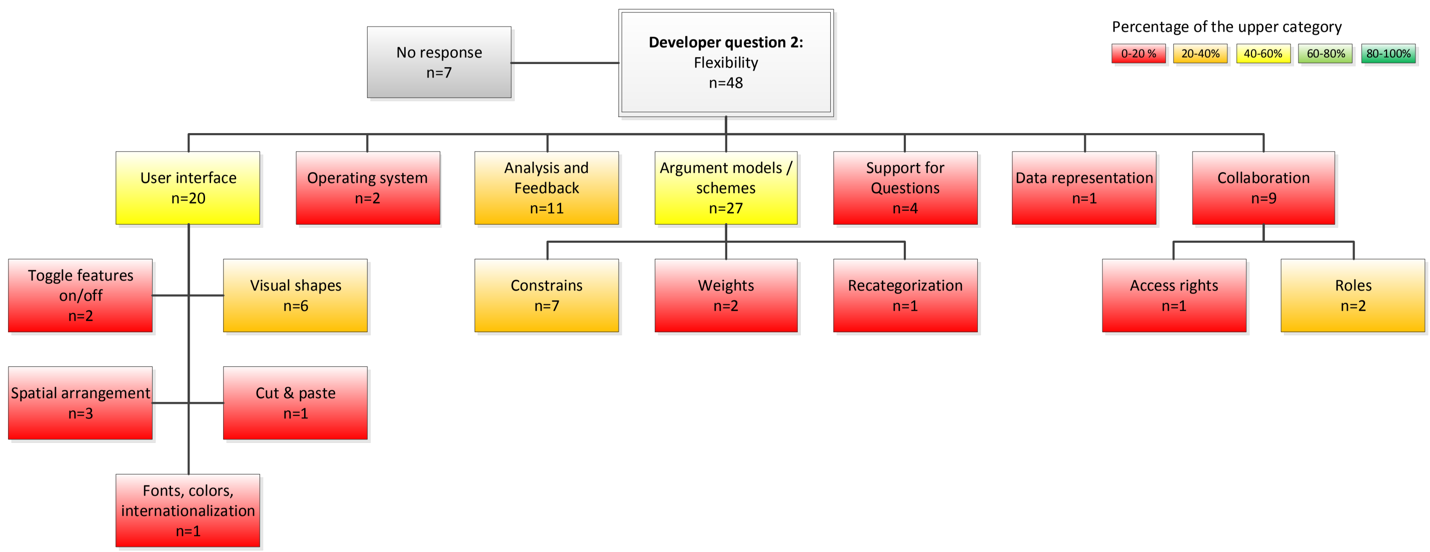


Figure C.8: Taxonomy of developer question 2: Can you briefly describe the types of flexibility and/or configuration you provided in the design of your argumentation system (or systems)?

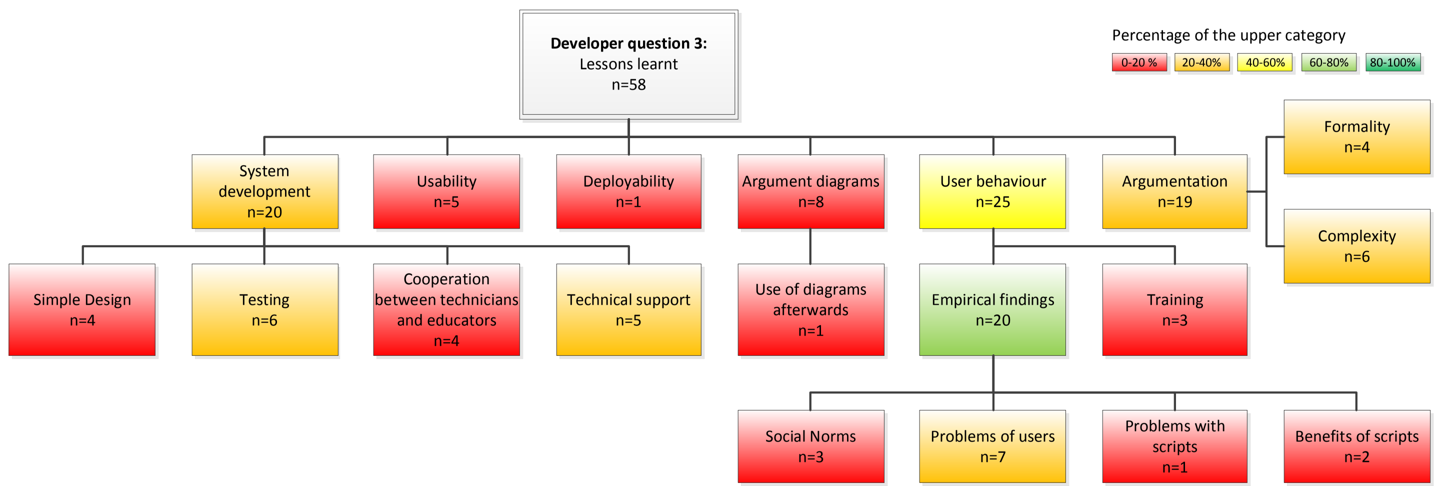


Figure C.9: Taxonomy of developer question 3: What would you say was the most important lesson (or lessons) learned in designing, developing and/or testing your argumentation system?

D Overview of Rich Internet Application Frameworks

In the following, a set of RIAs will be presented. For each framework, there will be a *spread* reported. This should only be used as indicator since these numbers are highly dependent on the measurement. The descriptions of the technologies are based in the information available on their project home websites. Whenever version numbers are reported they have been retrieved in July 2011.

D.1 Adobe Flash / Flex

Websites http://www.adobe.com/flashplatform/, http://www.adobe.com/products/flex/

Categorization Client-based execution / Installation of browser plug-in required

Description Adobe Flash was originally developed as tool for web designers and graphic artists. In order to use Flash for the development of complete applications, Flex evolved. Thus, Flex combines the graphical capabilities of Flash with an XML based markup language called MXML. It is used in combination with Action-Script to actually develop RIAs. Currently, Flex is in version 4.5.

Spread The Flash plug-in is installed on most desktop computers[2]. In the mobile device sector, Flash is less spreaded. As a matter of fact, some systems such as the iOS based devices do not support Flash at all.

Example applications YouTube, browser games, http://flex.org/showcase.

D.2 Apache Wicket

Website http://wicket.apache.org/

Categorization Server-based execution / No additional client-side installation required

Description Apache Wicket is a Java based framework that was designed to strictly seperate GUI and logic programming. Whereas the GUI is pure HTML and CSS, the logic uses Java and is located on the server's side. This way, there are no additional installation requirements on the client's side. Apache Wicket was introduced in 2004. Currently it is available in version 1.4.17.

Spread All web browser support HTML.

Example applications See http://wicketstuff.org/wicket14/ for example applications.

[2]According to the July 21, 2011 statistics from http://www.riastats.com/ 96.64% of the computers have Adobe Flash 9 or higher installed. Adobe claims even a spread of 99% (http://www.adobe.com/products/player_census/flashplayer/)

D.3 Google Web Toolkit

Website http://code.google.com/webtoolkit/

Categorization Client-based execution / No additional client-side installation required

Description Introduced in May 2006, the Google Web Toolkit (GWT) was designed to simplify browser independent development of web applications. The current version is 2.3. The development takes place in Java, which is compiled afterwards into JavaScript. The latter is executable in all modern web browsers without any additional plugins. To enable communication between server and client, GWT makes use of asynchronous remote procedure calls. By means of GWT the developer is not required to be an expert in web techniques to be able to create web applications.

Spread All modern browser support JavaScript.

Example applications Google Wave, Adwords, http://gwtgallery.appspot.com/.

D.4 JavaFX

Website http://javafx.com/

Categorization Client-based execution / Installation of local runtime environment required

Description JavaFX is the web pendant to Java. Similar to the latter, JavaFX applications will be compiled into platform independent Java bytecode. In the current version (1.3.1), it enriches Java by means of a scripting language called JavaFX Script in order to provide additional features that are typically used in web application. Further, additional user interface elements can be build using web techniques such as Cascading Style Sheets (CSS). However, in the next version (2.0) the scripting language JavaFX script will be ported to Java to avoid the need to learn an additional language on the developer's side.

Spread The Java Runtime Environment, which is required to execute JavaFX applications, is installed on a majority of end-user computers[3].

Example applications http://jfx.wikia.com/wiki/Demos

[3]According to the July 21, 2011 statistics from http://www.riastats.com/ 63.95% of the computers have Java 1.4 or higher installed.

D.5 Microsoft Silverlight

Website http://www.microsoft.com/silverlight/

Categorization Client-based execution / Installation of browser plug-in required

Description Microsoft Silverlight is an implementation of the .NET Framework for the web. That is, it is a sub set of the general desktop version enriched with multimedia capabilities especially designed to support web specific needs. The current version is 4. The development of Silverlight applications can be done by means of typical .NET languages such as C# or Visual Basic .NET.

Spread Microsoft Silverlight is installed on many end-user computers[4]. However, Microsoft Silverlight is based on the .NET Framework, which is available for Windows and MacOS only. A Linux pendant called Moonlight is supported by not directly developed by Microsoft. On the mobile device sector Microsoft Silverlight is only supported by devices which are running Windows Phone. In addition, Silverlight supports not all web browsers. Opera, for instance, is not directly supported.

Example applications http://www.silverlight.net/showcase

D.6 OpenLaszlo

Website http://www.openlaszlo.org/

Categorization Client-based execution / Installation of browser plug-in required OR no additional client-side installation required

Description OpenLaszlo evolved from the proprietary Laszlo Presentation Server. The first open source version has been introduced in October 2007. The current version is 4.9. In OpenLaszlo applications are developed by means of the XML based language LZX. The resulting source files were initially translated via Java on the server side into Flash bytecode. Since version 4.7 is is possible to use either Flash or DHTML as output format, which makes the framework more flexible since Flash is no longer essential.

Spread For the Flash condition, see spread of Adobe Flash. The DHTML condition is supported by all modern web browsers.

Example applications See http://openlaszlo.org/node/409 for an overview of showcases.

[4] According to the July 21, 2011 statistics from http://www.riastats.com/ 74.67% of the computers have Silverlight 3 or higher installed.

E Study Materials

If not mentioned seperately, the material chunks were taken from http://de.wikipedia.org. For the purpose of the study, it was not essential if all claims were valid. Thus, the use of Wikipedia content was adequate for the study. The references were not provided to the participants to avoid an extended internet search in the source.

E.1 Automobilverkehr der Zukunft - Brennstoffzelle, Elektrizität oder Biokraftstoff?

Aufgabenstellung

Die Automobilindustrie wird breitflächiger. Während Kraftfahrzeuge bisher überwiegend mittels Benzin, Diesel oder Erdgas angetrieben werden, ist man derzeit auf der Suche nach zukunftsträchtigen Technologien. Einige dieser Technologien sind die Brennstoffzelle, das Elektroauto sowie die Nutzung von Biokraftstoffen mit konventionellen Antrieben.

Diskutieren Sie auf Basis der gegebenen Materialien welches der drei Konzepte - Brennstoffzelle, Elektroauto oder Biokraftstoff - das größte Potenzial hat. Verdeutlichen Sie hierbei die Vor- und Nachteile der jeweiligen Konzepte und begründen Sie ihre Annahmen und Schlussfolgerungen mit den gegebenen Fakten.

Wichtige Anmerkungen

- Ein Besuch von anderen Webseiten als dem Softwaresystem ist zu keiner Zeit gestattet
- Der Gebrauch von Telefonen ist verboten

Bitte beachten Sie zudem folgende Punkte für Gruppenarbeiten:

- Die Kommunikation mit den Gruppenpartnern ist ausschließlich über das im Softwaresystem bereitgestellte Chatfenster erlaubt, d. h. es darf nicht direkt kommuniziert werden
- Verwenden Sie ausschließlich die gegebenen anonymen Decknamen

Bei Nichtbeachtung dieser Punkte erfolgt der sofortige Ausschluss aus dem Experiment.

Gemeinsame Materialien

Definition: Wattstunde (Einheitenzeichen: Wh)

Eine Maßeinheit der Arbeit und damit eine Energieeinheit. Eine Wattstunde entspricht der Energie, welche eine Maschine mit einer Leistung von einem Watt in einer Stunde aufnimmt oder abgibt. Im Alltag gebräuchlich und verbreitet ist die Kilowattstunde (kWh), das Tausendfache der Wattstunde. In ihr werden vor allem Strom-, aber auch Heizwärmekosten abgerechnet.

Leistungsbeispiele

Mit der Energiemenge 1 kWh kann man zum Beispiel:

- 50 Stunden am Laptop arbeiten (bei einer Leistung von 20 Watt)
- Sieben Stunden fernsehen (bei einer Leistung von ca. 142 Watt)
- Fünf Stunden am Computer arbeiten (bei einer Leistung von ca. 200 Watt)

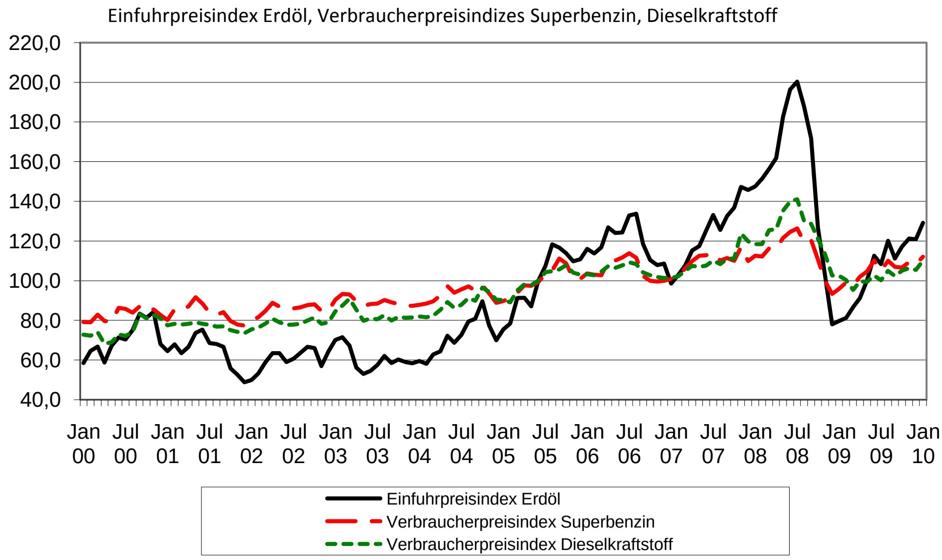


Abbildung E.10: Preisindizes 2000-2010 (Quelle: Statistisches Bundesamt Deutschland)

Definition: Wirkungsgrad

Verhältnis von abgegebener Leistung zu zugeführter Leistung. Die dabei entstehende Differenz von zugeführter und abgegebener Leistung bezeichnet man als Verlustleistung.

Fahrzeugkonzept	**Wirkungsgrad**
Auto mit Verbrennungsmotor (fossile Kraftstoffe oder Biokraftstoffe)	ca. 15% - 30%
Brennstoffzellenfahrzeug	ca. 35% - 50%
Elektrofahrzeug	ca. 80% - 95%

Tabelle E.2: Ungefährer Wirkungsgrad verschiedener Antriebe

Position 1: Der Elektroantrieb

Gustav Trouvé stellte 1881 auf der Internationalen Elektrizitätsausstellung ein elektrisch betriebenes, dreirädriges Automobil vor. Am 29. April 1882 führte Werner Siemens in Halensee bei Berlin einen elektrisch angetriebenen Kutschenwagen auf einer 540m langen Versuchsstrecke vor.

Ein Tesla Roadster (Serienfahrzeug, gebaut seit 2008, 252 PS, Basisausstattung: 109.000$) mit einer Speicherkapazität von 55 kWh ist in der Lage mit einer Ladung zwischen 90 und 501 Kilometer zurückzulegen.

Treibstoff	**Energiedichte [kWh/kg]**	**mittlerer Wirkungsgrad**	**Gesamtmasse des Energiespeichers in kg für 100 kWh nutzbare Energie***
Strom aus Doppelschicht-Kondensator	0,005	95%	16200
Strom aus Bleiakkumulator	0,03	95%	2700
Strom aus Lithium-Ionen-Akkumulator	0,13	95%	623
Dieselkraftstoff	11,8	25%	36 (+50 Tankbehälter)
Superbenzin	12,0	15%	59 (+50 Tankbehälter)
Flüssiger Wasserstoff	33,3	38%	8,3 (+600 Tankbehälter)

**Bei Akkus ist die Masse des Behälters im Wert der Energiedichte bereits enthalten. Bei Diesel/Benzin/Wasserstoff muss er addiert werden*

Tabelle E.3: Treibstoffe im Vergleich: Energiedichte pro Kilogramm

Jahr	**€ je kWh**	**absolut in €/kWh**	**relativ in %**
1996	0,1320		
1997	0,1270	−0,0050	−3,8
1998	0,1256	−0,0014	−1,1
1999	0,1277	+0,0021	+1,7
2000	0,1191	−0,0086	−6,7
2001	0,1220	+0,0029	+2,4
2002	0,1261	+0,0041	+3,4
2003	0,1267	+0,0006	+0,5
2004	0,1259	−0,0008	−0,6
2005	0,1334	+0,0075	+6,0
2006	0,1374	+0,0040	+3,0
2007	0,1433	+0,0059	+4,3
2008	0,1299	−0,0134	−9,3
2009	0,1401	+0,0102	+7,85

Am Endkundenpreis für Strom sind ca. 40% Steuern und Abgaben enthalten.

Tabelle E.4: Durchschnittsstrompreis (Deutschland, ohne Steuern und Abgaben)

Elektromotoren sind einfacher aufgebaut und besitzen erheblich weniger bewegliche Teile als Verbrennungsmotoren. Ölwechsel sind nicht notwendig. Elektrisch betriebene Kraftfahrzeuge arbeiten, was den Antriebsteil angeht, in der Regel sehr wartungsarm. Seit einiger Zeit experimentieren diverse Hersteller mit Radnabenmotoren. An den Antriebsachsen sind dabei die Räder mit je einem eigenen Motor innerhalb des Rades ausgestattet. Bei dieser Art des Antriebes entfallen zum Beispiel der „Motorraum" und viele Teile des konventionellen Antriebsstranges und vereinfachen so den Aufbau.

Im Teillastbetrieb wirken sich Wirkungsgradunterschiede besonders stark aus.

Automobile im Stadtverkehr fahren fast immer mit Teillast.

Ein Elektromotor verbraucht während des Fahrzeugstillstands keine Energie.

Elektrofahrzeuge besitzen häufig die Fähigkeit, beim Bremsen durch Nutzbremsung einen Teil der Antriebsenergie zurückzugewinnen.

Der Wirkungsgrad im (Elektro-)Fahrzeug blendet die Verluste bei der Bereitstellung des Stroms aus. Die Wirkungsgrade der Stromkraftwerke liegen zwischen 30% und 60%. Dazu kommen Leitungsverluste. Bezieht man dies in die Betrachtung mit ein, ist die Klimabilanz von Elektroautos in etwa gleich wie bei Autos mit Verbrennungsmotor. Unter Verwendung von „Ökostrom" schneiden Elektroautos jedoch deutlich besser ab.

Elektroautos verlagern die Emissionen für ihren Betrieb vom Fahrzeug weg zu den Kraftwerken, in denen der Strom für ihren Betrieb produziert wird.

Elektroautos arbeiten zumeist sehr leise, was daran liegt, dass sie keine lauten explosionsartigen Verbrennungen zur mechanischen Energieerzeugung, wie z. B. Dieselmotoren, nutzen. Feinstaub-Emissionen durch Reifenabrieb und Bremsvorgänge bleiben erhalten.

Das Aufladen eines Akkus braucht derzeit mehrere Stunden.

Die vollständige Umstellung aller PKW in Deutschland auf Elektrofahrzeuge würde den Strombedarf Deutschlands um bis zu 60 Gigawatt steigern. (1 Gigawatt (GW) sind eine Milliarde Watt; 1 GW entspricht der Leistung eines typischen Kernkraftwerks; 122 GW entspricht der Leistung aller in Deutschland installierten Kraftwerke (Stand 1997)).

Die Lebensdauer von (kostengünstigen) Bleiakkumulatoren beträgt zwischen 5.000 und 50.000 km. Bleiakkumulatoren werden seit Jahren im Auto eingesetzt. Die wohl bekannteste Anwendung ist die Starterbatterie für Kraftfahrzeuge.

Ein Doppelschicht-Kondensator ist ein Energiespeicher für elektrische Energie und dem Akkumulator in praktisch allen Kennwerten außer der Energiedichte weit überlegen.

Ein Lithium-Ionen Akkusatz mit 10 kWh Kapazität kostete 2008 etwa 5.000 €. Die Lebensdauer ist beschränkt. Zurzeit gilt die Faustregel, dass ein Lithium-Ionen-Akku nach ca. drei Jahren mehr als 50 % seiner Kapazität eingebüßt hat. Dieser Wert ist abhängig von diversen Faktoren, insbesondere der richtigen Lagerung (Temperatur, Ladezustand).

„Hybridautos haben zusätzlich zum klassischen Verbrennungsmotor einen Akku an Bord. Wenn der leer ist, springt der Benziner an. Eine Variante sind sogenannte Mild-Hybrid-Systeme, bei denen der Stromantrieb nur parallel unterstützend läuft, um den Benzinverbrauch zu reduzieren. Der Akku wird in der Regel durch Bremskraftrückgewinnung und einen Dynamo geladen. Zukünftige Hybridfahrzeuge sollen aber auch an der Steckdose aufladbar sein.“ (Spiegel.de)

Position 2: Das Brennstoffzellenfahrzeug

Brennstoffzellenfahrzeuge sind Transportmittel mit Elektroantrieb, bei denen die benötigte elektrische Energie aus Wasserstoff durch eine Brennstoffzelle erzeugt wird.

Im Juni 2008 ging mit dem Honda FCX das erste Fahrzeug in (Klein)-Serienproduktion.

Wasserstoff wird entweder als Gas oder in tiefkalter flüssiger Form (-253°C) in Tanks gelagert.

Die Produktion eines Wasserstofftanks ist aufwändig hinsichtlich Dichtheit, Wärmeisolierung und Sicherheit.

Wasserstoff ist ein farbloses, geschmacks- und geruchsloses, ungiftiges Gas. In der Natur findet man es praktisch nicht in freier Form. Es liegt ausschließlich in gebundener Form, z.B. als Wasser, in Kohlenwasserstoffen (Erdöl, Erdgas, Kohle, Biomasse) oder in anderen organischen Verbindungen vor.

Die Gewinnung von Wasserstoff erfolgt entweder aus fossilen Energieträgern oder durch Elektrolyse von Wasser. Unter Elektrolyse versteht man die Aufspaltung einer chemischen Verbindung unter Einwirkung des elektrischen Stroms. Folglich erfordert auch die Aufspaltung von Wasser in Wasserstoff Energie.

Kraftstoff	**Energiemenge (kWh) pro** m^3 [höher ist besser]
Benzin	8760 kWh/m^3
Flüssiggas	6966 kWh/m^3
Erdgas (unter Hochdruck)	2580 kWh/m^3
Wasserstoff (flüssig unterhalb -253 °C	2360 kWh/m^3
Wasserstoffgas (unter Hochdruck)	530 kWh/m^3
Wasserstoffgas (unter Normaldruck)	3 kWh/m^3

Tabelle E.5: Kraftstoffe im Vergleich: Energiemenge pro Kubikmeter

Eine Brennstoffzelle kann chemisch gebundene Energie (bspw. in Form von Wasserstoff) mit einem Wirkungsgrad von ca. 35-50% direkt in elektrische Energie umwandeln. Der so gewonnene Strom wird in Elektromotoren, die oft ohne Getriebe direkt an zwei oder vier Rädern montiert sind, in Bewegungsenergie umgewandelt.

Die Umwandlung von chemischer in mechanische Energie verursacht nur geringe Geräuschemissionen.

„*Mercedes-Benz: Brennstoffzellenfahrzeug in Kleinserie*

Auf verschiedenen Wegen will Daimler zu einer emissionsfreien Mobilität gelangen. Neben der Optimierung von Verbrennungsmotoren und der Entwicklung von Hybrid- und Elektrofahrzeugen basteln die Schwaben bereits seit vielen Jahren an einem Brennstoffzellenfahrzeug. Nun startet Ende 2009 die Serienproduktion der B-Klasse F-Cell, zunächst in einer Kleinserie von 200 Einheiten. Wegen der nicht bezifferten, aber immens hohen Herstellungskosten werden die Brennstoffzellenfahrzeuge nicht verkauft, sondern können geleast werden. Dies obliegt aufgrund eines weltweit fehlenden Wasserstoff-Tankstellennetzes allerdings nur Firmen, die die Fahrzeuge dank ihrer unternehmerischen Ausrichtung auch betanken können.

Bei der B-Klasse F-Cell treibt die mit Wasserstoff zu betankende Brennstoffzelle einen Elektromotor mit einer Leistung von 100 kW/136 PS an. Das Aggregat entwickelt ein maximales Drehmoment von 320 Nm. In der Praxis lässt sich das Fahrzeug damit so gut fahren wie ein Auto mit gleich starkem Verbrennungsmotor und Doppelkupplungsgetriebe. Ruckfrei beschleunigt der Mercedes und lässt sich agil über die Landstraße bewegen - und mit gutem Gewissen, denn aus dem Auspuff strömt nur noch etwas Wasserdampf. Angenehm fällt dabei auch die geringe Geräuschemission auf: nur noch das Reifenabrollgeräusch und der sich an der Karosserie brechende Wind ist zu hören.“ (Spiegel.de, 15.07.2009)

In Deutschland befinden sich derzeit sieben öffentliche Wasserstoff-Tankstellen. Dort kostet ein Kilo Wasserstoff zwischen acht und neun Euro. Die Mercedes B-Klasse F-Cell verbraucht etwa ein Kilo auf 100 Kilometer. Bei Mercedes rechnet man künftig gar mit einem Preis von gut drei Euro pro Kilo Wasserstoff.

Position 3: Der Biokraftstoff

Biokraftstoffe sind flüssige oder gasförmige Kraftstoffe, die aus Biomasse hergestellt werden. Ausgangsstoffe für Biokraftstoffe sind nachwachsende Rohstoffe, wie z. B. Ölpflanzen, Getreide, Zuckerrüben oder Zuckerrohr, Wald- und Restholz, Holz aus Schnellwuchsplantage, spezielle Energiepflanzen und anderes.

Biokraftstoffe sind in der Lage fossile Kraftstoffe (Diesel, Benzin, Erdgas) zu ersetzen. Dazu müssen teilweise Motoren und/oder Kraftstoffsysteme angepasst werden.

Es werden verschiedene Arten von Biokraftstoffen unterschieden: Pflanzenöl-Kraftstoff, Biodiesel, Bioethanol, Biomethan und synthetische Biokraftstoffe.

Pflanzenöl-Kraftstoff besteht aus unbehandeltem oder raffiniertem Pflanzenöl in Reinform. In Deutschland ist der Grundstoff in der Regel Rapsöl. In den chemischen Eigenschaften unterscheidet es sich vom Dieselkraftstoff, weshalb eine Anpassung der Motoren an diesen Kraftstoff erforderlich ist.

Biodiesel wird aus Pflanzenölen hergestellt. Mit Biodiesel kann Dieselkraftstoff substituiert (ersetzt) werden. In Deutschland ist der Grundstoff meist Rapsöl.

Bioethanol wird durch Vergärung biogener Rohstoffe und anschließende Destillation hergestellt. Mit Bioethanol kann Benzin substituiert werden. In Deutschland wird für die Herstellung von Bioethanol meist Getreide, Mais und Zuckerrübe verwendet. In Brasilien deckt Ethanol aus Zuckerrohr einen großen Teil des nationalen Treibstoffbedarfs. Die chemischen Eigenschaften unterscheiden sich vom Benzin, weshalb eine Anpassung der Fahrzeugmotoren erforderlich ist.

Biomethan wird aus dem Vorprodukt Biogas hergestellt. Für die Erzeugung von Biogas kommen in der Regel Energiepflanzen, Gülle und/oder organische Reststoffe als Gärsubstrate zum Einsatz. Mit Biomethan kann Benzin oder Erdgas substituiert werden. Fahrzeuge, die für den Einsatz von reinem oder bivalentem Erdgasbetrieb umgerüstet sind, können mit Biomethan betrieben werden.

Synthetische Kraftstoffe können aus verschiedenen organischen Rohstoffen hergestellt werden. Sie können auf die Erfordernisse moderner Motoren zugeschnitten werden und beispielsweise Dieselkraftstoff ersetzen. Derzeit befinden sich diese Kraftstoffe noch im Entwicklungsstadium und sind noch nicht auf dem Markt erhältlich.

Einige Biokraftstoffe können auch regional in dezentralen, kleinen Produktionsanlagen wirtschaftlich hergestellt werden, wie z. B. Pflanzenöl und Bioethanol in landwirtschaftlichen Betrieben bzw. in kleinen Alkoholbrennereien. Aber auch in Großanlagen werden sie gewonnen. Anlagen zur Produktion von Biodiesel und synthetischem Kraftstoff dagegen erfordern Errichtung und Betrieb komplexer, größerer, überregionaler Produktionseinheiten.

Die Besteuerung von Biokraftstoffen ist in Deutschland im §50 Energiesteuergesetz geregelt. Als Beimischung in fossilen Kraftstoffen unterliegen Biokraftstoffe dem vollen Steuersatz für Kraftstoffe. Für reine Biokraftstoffe dagegen ist die Energiesteuer reduziert. Für Bioethanol und Biomethan gilt eine komplette Steuerermäßigung. Für Pflanzenöl-Kraftstoff und Biodiesel muss ein Steueranteil gezahlt werden, der jährlich ansteigt, bis der volle Steuersatz für fossile Kraftstoffe erreicht ist.

„Volle Tanks, leere Teller

[...] New York/Mexiko Stadt - In den USA erleben Biotreibstoffe einen ungekannten Boom. Der Absatz für Ethanol-Sprit schießt in die Höhe und beschert den Produzenten satte Gewinne. Die negativen Folgen spüren vor allem die Armen im südlichen Nachbarland Mexiko: Hier wird der Mais knapp - und deshalb immer teurer. Mais ist seit Menschengedenken das wichtigste Grundnahrungsmittel der Mexikaner, die ihn in großen Mengen zu Tortilla-Fladen backen. Die uralte Feldfrucht ist aber auch Grundlage für Ethanol. Wegen der starken Nachfrage in den USA haben die Maispreise an den internationalen Rohstoffmärkten den höchsten Stand seit zehn Jahren erreicht. Die unkalkulierbaren Schwankungen auf dem Markt für fossile Rohstoffe lassen die energiehungrige Supermacht vermehrt auf erneuerbare Energien setzen. Der Boom lenkt die Handelsströme: Die Maisernte fließt dahin, wo das Geld ist, und schreibt dabei eine Geschichte über Gewinner und Verlierer im globalen Wettbewerb. In Mexiko-Stadt verdoppelte sich der Kilopreis für Tortilla innerhalb weniger Wochen von umgerechnet 40 auf 75 Euro-Cent. [...]

Die Zahlen des US-Landwirtschaftsministeriums sprechen für sich: Vor sechs Jahren gab es in den USA gut 50 Ethanol-Produzenten mit einer Jahresproduktion von weniger als acht Milliarden Litern. Inzwischen erzeugen über 100 Firmen mehr als 18 Milliarden Liter. Derzeit sind 70 Fabriken mit einer Kapazität von zusätzlich acht Milliarden Litern im Bau. Inzwischen fließen bereits 20 Prozent der US-Maisernte in die Ethanolgewinnung. Im Jahr 2000 waren es noch sechs Prozent.

In Mexiko hingegen dient der Mais noch seiner klassischen Bestimmung: Er soll Mägen füllen, nicht Autotanks. Angesichts des öffentlichen Unmuts über die Tortilla-Krise hat sich inzwischen Präsident Felipe Calderón eingeschaltet. An einem runden Tisch setzte er sich am Mittwoch mit Agrarunternehmern zusammen und handelte eine staatlich bestimmte Obergrenze von umgerechnet 60 Euro-Cent pro Kilo Tortilla aus. [...]“ (Spiegel.de, 23.01.2007)

Biokraftstoffe kommen als Reinkraftstoff und als Beimischungen zu fossilen Kraftstoffen zum Einsatz. Mit der EU-Richtlinie 2003/30/EG wurden Beimischungen von 2% bis 2005, 2,75% bis 2006 und 5,75% bis 2010 gefordert. Wegen der mangelnden Umsetzung wurde in der EU-Richtlinie 2009/28/EG (Biokraftstoffrichtlinie) ein verbindlicher Wert von 10% bis 2020 festgelegt. Gemäß dem Biokraftstoffquotengesetz müssen in Deutschland derzeit (2009) fossilen Kraftstoffen 5,25% Biokraftstoffe beigemischt werden, bezogen auf den Energiegehalt des Kraftstoffs.

Gegenwärtig werden 5% der globalen Getreideernte zur Herstellung von Biokraftstoffen genutzt. Von der europäischen Getreideernte werden 1,6% für Biokraftstoffe genutzt. Der überwiegende Teil (58%) wird für Viehfutter verwendet.

E.2 Der Energiemix in Deutschland im Jahr 2030

Aufgabenstellung

Die Energieversorgung in Deutschland befindet sich Wandel. Während in der Vergangenheit hauptsächlich fossile Energieträger wie Erdöl, Kohle oder Erdgas zur Energiegewinnung genutzt wurden, sind in den letzten Jahren vermehrt alternative Konzepte zur Stromerzeugung mittels erneuerbaren Energien wie Wind, Wasser- und Sonnenenergie im Focus der Öffentlichkeit.

Diskutieren Sie auf Basis der gegebenen Materialien wie der Energiemix in Deutschland *im Jahre 2030* aussehen könnte. Verdeutlichen Sie hierbei die Vor- und Nachteile der jeweiligen Konzepte und begründen Sie ihre Annahmen und Schlussfolgerungen mit den gegebenen Fakten.

Wichtige Anmerkungen

- Ein Besuch von anderen Webseiten als dem Softwaresystem ist zu keiner Zeit gestattet
- Der Gebrauch von Telefonen ist verboten

Bitte beachten Sie zudem folgende Punkte für Gruppenarbeiten:

- Die Kommunikation mit den Gruppenpartnern ist ausschließlich über das im Softwaresystem bereitgestellte Chatfenster erlaubt, d. h. es darf nicht direkt kommuniziert werden
- Verwenden Sie ausschließlich die gegebenen anonymen Decknamen

Bei Nichtbeachtung dieser Punkte erfolgt der sofortige Ausschluss aus dem Experiment.

Gemeinsame Materialien

Definition: CO_2

CO_2 steht für Kohlenstoffdioxid und ist eine chemische Verbindung aus Kohlenstoff und Sauerstoff. Es entsteht sowohl bei der vollständigen Verbrennung von kohlenstoffhaltigen Substanzen unter ausreichender Sauerstoffzufuhr als auch im Organismus von Lebewesen. CO_2 ist natürlicher Bestandteil der Luft (mittlere Konzentration von 0,038%), wird jedoch für die globale Erwärmung, d.h. für den Klimawandel, mitverantwortlich gemacht.

Kraftwerksart	**CO_2-Emissionen pro kWh in Gramm**	**Anteil der gesamten elektr. Energie in Deutschland**	**Anteil an der CO_2-Erzeugung aller Kraftwerke in Deutschland**
Wasserkraft	4-13	4,3%	0,06%
Windenergie	8-16	6,2%	0,12%
Kernkraftwerk	16-23	22%	0,7%
Photovoltaik	20-100	0,5%	0,1%
Erdgas	410-430	11,7%	8,1%
Erdöl	890	1,3%	1,9%
Steinkohle	790-1080	22,8%	35,3%
Braunkohle	980-1230	24,5%	44,9%
Andere (Müll, Biomasse, ...)	800 (geschätzt)	6,7%	8,9%

Anmerkungen: Die CO_2-Emissionen enthalten die jeweiligen CO_2-Werte, die durch Herstellung, Betrieb und Abriss entstehen.

Tabelle E.6: Anteil und CO_2 Ausstoß von unterschiedlichen Kraftwerksarten

Definition: Wattstunde (Einheitenzeichen: Wh)

Eine Maßeinheit der Arbeit und damit eine Energieeinheit. Eine Wattstunde entspricht der Energie, welche eine Maschine mit einer Leistung von einem Watt in einer Stunde aufnimmt oder abgibt. Im Alltag gebräuchlich und verbreitet ist die Kilowattstunde (kWh), das Tausendfache der Wattstunde. In ihr werden vor allem Strom-, aber auch Heizwärmekosten abgerechnet.

Leistungsbeispiele

Mit der Energiemenge 1 kWh kann man zum Beispiel:

- 50 Stunden am Laptop arbeiten (bei einer Leistung von 20 Watt)
- Sieben Stunden fernsehen (bei einer Leistung von ca. 142 Watt)
- Fünf Stunden am Computer arbeiten (bei einer Leistung von ca. 200 Watt)

Biokraftstoff	**Ertrag** / *ha*	**Kraft-stoff-äqui-valenz***	**Kraftstoff-äquivalent pro Fläche** (l/ha)	**Preis in €**	**Preis Kraftstoff-äquivalent in €**
Pflanzenöl (Rapsöl)	1590 l	0,96	1526 l	0,981€/l	1,022€/l
Biodiesel (Rapsmethylester)	1550 l	0,91	1411 l	1,079€/l	1,186€/l
Bioethanol (Weizen)	2760 l	0,65	1794 l	0,932€/l	1,331€/l
Biomethan	3540 kg	1,4	4956 l	0,93€/kg	0,66€/l
Synthetische Kraftstoffe	4030 l	0,97	3909 l	nicht am Markt	k. A.

**1 l Biokraftstoff bzw. 1 kg Biomethan entspricht dieser Menge konventionellen Kraftstoffs*

Tabelle E.7: Vergleich von Biokraftstoffen in Deutschland

Der Bedarf an Strom schwankt tageszeitabhängig stark. Insbesondere tagsüber kommt es zu Lastspitzen.

Position 1: Kernenergie

Um 1890 wurden erste Experimente zur Radioaktivität durchgeführt. Zuerst wurden diese Erkenntnisse für die militärische Forschung während des Zweiten Weltkrieges eingesetzt. 1954 wurde in Obninsk bei Moskau das erste kommerzielle Kernkraftwerk der Welt in Betrieb genommen. In Deutschland wurde 1957 der erste Forschungsreaktor in Betrieb genommen. 1961 folgte das erste deutsche Kernkraftwerk „Kahl" mit einer Leistung von 15 Megawatt (MW). In den 1960er Jahren wurden zahlreiche weitere Kernkraftwerke gebaut, wobei deren Leistung erheblich erhöht wurde. In den 1970er Jahren wurde insbesondere nach der ersten Ölkrise 1973 der Bau von Kernkraftwerken forciert. Die Leistung dieser Kraftwerke, wie etwa des Blocks B des Kernkraftwerks Biblis, lag bei 1,3 Gigawatt (GW). Das Kernkraftwerk Biblis ist auch heute noch im Betrieb.

Am 26. April 1986 ereignete sich die Katastrophe von Tschernobyl im Kernkraftwerk Tschernobyl als Folge einer Kernschmelze und Explosion im Kernreaktor Tschernobyl Block 4. Bei diesem Super-GAU wurden große Mengen von Radioaktivität freigesetzt. Es gilt als die schwerste nukleare Havarie und als eine der schlimmsten Umweltkatastrophen aller Zeiten. Als Ursache gelten grundlegende Mängel in der Konstruktion des Reaktors sowie Planungs- und Bedienungsfehler bei einem Versuch (Simulation eines totalen Stromausfalls).

Im Jahr 2000 wurde in Deutschland auf Druck der Bundesregierung der Ausstieg aus der kommerziellen Nutzung der Kernenergie bis etwa 2020 beschlossen. In diesem Rahmen wurden bis 2005 bereits zwei Kernkraftwerke vom Netz genommen, der Ausstiegsbeschluss ist jedoch politisch und gesellschaftlich weiter umstritten.

Weltweit sind 438 Kernkraftwerke mit einer Gesamtleistung von 372 GW in 31 Ländern in Betrieb. Weiterhin sind 42 neue Kernkraftwerke im Bau. (Stand 31.12.2008)

„Frankfurt am Main - In der Bundesregierung tobt der Streit um die Atomenergie: Die Union will Deutschlands Kernkraftwerke länger laufen lassen, die SPD-Führung ist dagegen und hält am Ausstieg fest. Nun mischen sich auch Verbraucherschützer in die Debatte ein - und schlagen sich auf die Seite der Atomgegner.

Sollten die deutschen Kernkraftwerke länger laufen, würde sich dies im Portemonnaie der Bundesbürger nur minimal bemerkbar machen. Das geht aus Berechnungen des Bundesverbands der Verbraucherzentralen (vzbv) hervor. Längere Atomlaufzeiten brächten eine Ersparnis von gerade einmal 50 Cent pro Monat, sagte der Energiefachmann des vzbv, Holger Krawinkel. Beim Austausch einer einzigen 60-Watt-Glühbirne durch eine Energiesparlampe könne man mehr Geld sparen. [. . .]" (Spiegel.de, 07.07.2008)

Bei der Kernspaltung entstehen radioaktive Isotope. Verbrauchte Brennelemente sind stark strahlend. Direkt nach dem Einsatz ist die Strahlung so stark, dass eine weitere Verarbeitung nicht möglich ist. Die Brennelemente werden aus diesem Grund für einige Monate, oder Jahre im Zwischenlager des Kernkraftwerks in einem Abklingbecken gelagert. Nach dieser Zeit sind kurzlebige Isotope weitgehend zerfallen. Es verbleiben jedoch langlebige Isotope, die die Brennelemente weiterhin hoch radioaktiv machen.

Abgebrannte, nicht wiederaufgearbeitete Brennelemente und radioaktiver Abfall aus Wiederaufbereitungs-anlagen werden in Lagerungsbehältern in Zwischenlagern so lange gelagert, bis die Radioaktivität so weit abgeklungen ist, dass eine Endlagerung möglich ist. Dies dauert einige Jahrzehnte.

Der radioaktive Abfall eines Kernkraftwerks weist auch nach Jahrzehnten eine hohe Radioaktivität auf, die nur langsam auf ein ungefährliches Maß absinkt. Je nachdem, was als ungefährlich eingestuft wird, sind dafür einige Tausend bis einige Hunderttausend Jahre erforderlich.

In Wiederaufarbeitungsanlagen können die in abgebrannten Brennelementen zu ca. 10% wiederverwendet werden. Die anderen 90% sind zum Teil hoch radioaktiver Abfall.

Kernkraftwerke sind ganzjährig nutzbar. Das heißt sie sind nicht etwa von Wetter oder Klima abhängig.

Staat	**Gesamterzeugung aller Kraftwerke in TWh**	Strommix g CO_2 pro kWh	Gesamt-CO_2 in 10^9 kg	Anzahl großer thermischer Kraftwerksblöcke	Anzahl der Kernkraftwerksblöcke
Deutschland	636,5	604	384	ca. 70	17
Frankreich	610,6	61	37	15	58

Tabelle E.8: CO_2 Ausstoß durch Erzeugung elektrischer Energie - Deutschland und Frankreich im Vergleich

Die Wiederaufarbeitungsanlage „La Hague“ in Frankreich leitet jährlich ca. 6,6 Mrd. Liter radioaktiv kontaminiertes Abwasser in den Ärmelkanal. Zudem wird radioaktiv kontaminierte Abluft über Europa freigesetzt.

Das Wissen über Nutzung der Kernkraft erhöht auch die Gefahr eines terroristischen Einsatzes von Nuklearwaffen.

Die Schäden eines Unfalls in einem Atomkraftwerk mit erheblicher Freisetzung von Radioaktivität bezifferte eine Studie der Prognos AG 1992 mit 5 bis 12 Billionen DM (ca. 2,5 bis 6 Billionen €), was dem drei- bis vierfachen des damaligen jährlichen deutschen Bruttosozialproduktes entspricht.

Kernkraftwerke erzeugen im laufenden Betrieb kein CO_2.

Deutsche Kernkraftwerke haben keine Haftpflichtversicherung für einen nicht beherrschbaren Störfall. Die Folgekosten eines solchen Super-GAUs müsste der Staat, d.h. der Steuerzahler, übernehmen. Diese Kosten sind nicht im Strompreis enthalten.

In Brennelementen von Kernreaktoren wird überwiegend Uran (95%) und Plutonium (1%) verwendet.

Für Atomkraftwerke hat im vergangenen Jahr (2008) eine Untersuchung des Bundesamtes für Strahlenschutz bewiesen: Das Risiko für Kindern an Leukämie (Blutkrebs) zu erkranken nimmt zu, je näher sie an einem Kernkraftwerk wohnen.

Die weltweite Verfügbarkeit von Uran beträgt auch unter Berücksichtigung des weltweiten Kernenergie-ausbaues und den jetzigen Uranpreisen mehr als 100 Jahre. Mit höheren zulässigen Kosten für die Gewinnung des Urans reicht es für mehr als 1000 Jahre.

Die weltweit größten Uranreserven weltweit befinden sich in Australien und Kanada.

In Deutschland findet offiziell kein Uranabbau mehr statt.

Der mittlere Nettostrompreis für Haushalte mittlerer Größe liegt im EU-27 Durchschnitt bei 12,11 ct/kWh, in Frankreich bei 9,14 ct/kWh.

Position 2: Erneuerbare Energien (Wind-, Wasser-, Solarenergie)

Als Erneuerbare Energien bezeichnet man Energie aus Quellen, die sich entweder kurzfristig von selbst erneuern oder deren Nutzung nicht zur Erschöpfung der Quelle beiträgt.

In Deutschland werden Erneuerbare Energien mit unterschiedlichen Maßnahmen gefördert. Im Jahr 2000 wurde das Erneuerbare-Energien-Gesetz (EEG) erlassen. Es regelt, dass von privaten Anbietern erzeugter Strom aus Erneuerbaren Energien von Netzbetreibern zu Mindestpreise abgenommen werden muss.

Mit der EU-Richtlinie zu den Erneuerbaren Energien vom 23. April 2009 (2009/28/EG) wird den Mitgliedsstaaten der Europäischen Union der Erlass von Gesetzen vorgeschrieben, die die Verwendung der erneuerbaren Energien fördern, damit bis 2020 ein Gesamtanteil dieser Energien am Energiegesamtverbrauch innerhalb der EU von 20% erreicht wird.

Funktionsweise von Wasserkraftwerken: Durch eine Stauanlage wird Wasser im Stauraum auf möglichst hohem potentiellem Niveau zurückgehalten. Die Energie der Bewegung des abfließenden Wassers wird auf eine Wasserturbine oder ein Wasserrad übertragen, wodurch dieses in Drehbewegung mit hohem Drehmoment versetzt wird. Dieses wiederum wird direkt oder über ein Getriebe an die Welle des Generators weitergeleitet, der die mechanische Energie in elektrischen Strom umwandelt.

Die auf die Erde eingestrahlte Sonnenenergie entspricht etwa dem Zehntausendfachen des aktuellen menschlichen Energiebedarfs.

Laut einer repräsentativen Forsa-Umfrage 2010 zur Akzeptanz Erneuerbarer Energien halten 95% der Deutschen den Ausbau von erneuerbaren Energien für wichtig oder sehr wichtig. 78% würden ihren Strom am liebsten aus erneuerbaren Energiequellen beziehen (im Vergleich zu 9% aus Erdgas, 6% aus Atomkraft, 3% aus Kohle).

Während fossile und atomare Energieträger immer teurer werden, sind die Kosten für Erneuerbare Energien in den letzten 15 Jahren im Schnitt um etwa die Hälfte gesunken. Bis 2020 strebt die Branche eine weitere Kostensenkung von 40% an, ermöglicht durch Massenfertigung und Technologiefortschritte.

Der Wandel von der konventionellen Energiebereitstellung zu erneuerbaren Energien verändert die Struktur der Energiewirtschaft massiv. Statt der Stromerzeugung in Großkraftwerken mit zum Teil mehr als 1000 Megawatt Leistung (Kern-, Braunkohle- und Steinkohlekraftwerke), nimmt die Erzeugung in Kleinanlagen mit wenigen kW (z. B. Photovoltaik / Solaranlagen) bis wenige MW (kleinere Windparks) zu.

Die Stromerzeugung aus Solar- und Windenergie unterliegt starken Schwankungen.

Wasserkraftwerke haben eine hohe Lebenszeit von bis zu 100 Jahren.

Die Herstellung von Photovoltaikanlagen (Solaranlagen) ist relativ energieaufwändig, sodass die energetische Amortisationszeit in Deutschland etwa sechs Jahre beträgt.

Die Lebenszeit von Solaranlagen beträgt 20 bis 30 Jahre.

Laut NABU sterben in Deutschland jährlich etwa eintausend Vögel durch Kollision mit einem Windrad, was ca. 0,5 Vögeln pro Anlage und Jahr entspricht.

Laut dem deutschen Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (BMU) hat sich die Zahl der Beschäftigten im Wirtschaftszweig erneuerbare Energien von 2004 (rund 160.500) bis 2008 (278.000) um ca. 73% erhöht. Nach Studien des BMU könnten bis zum Jahre 2020 über 400.000 Menschen in Deutschland im Bereich erneuerbare Energien beschäftigt sein.

Energiequelle	**1995**	**2000**	**2004**	**2005**	**2006**	**2007**	**2008**
Wasserkraft	77	92	76	77	78	76	75
Windenergie	6	35	92	95	110	143	145
Photovoltaik (Solarenergie)	0,03	0,3	2	4	7	11	14

**1 Petajoule entspricht* 10^{15} *Joule. 1 Joule entspricht 1 Wattsekunde, d.h. der Energie, die benötigt wird um für die Dauer von einer Sekunde eine Leistung von einem Watt aufzubringen.*

Tabelle E.9: Erzeugte Energien in Deutschland in Petajoule*

Windanlagen liefern im Winter durchschnittlich mehr Strom, als im Sommer.

Solaranlagen liefern nachts und im Winter weniger bis gar keinen Strom.

Bei Wasserkraftwerken kann die Energieumwandlung mehrere Wochen bis Monate aufgeschoben werden. Photovoltaik- und Windenergieanlagen können abgeschaltet und innerhalb weniger Minuten wieder in Betrieb genommen werden.

„[...] Es klingt wie aus einem Roman von Jules Verne: Afrika soll das Kraftwerk Europas werden. Die Pläne des Versicherungskonzerns Münchener Rück, die Energieversorgung auf eine völlig neue Grundlage zu stellen, elektrisiert seit Tagen die gesamte Strombranche. Ein Netz von solarthermischen Kraftwerken in der nordafrikanischen Wüste könnte bald sauberen und erstmals auch speicherbaren Solarstrom produzieren - und über leistungsfähige Gleichstromkabel nach Europa transportieren. „Es klingt bestechend“, gab Hermann Scheer zu, Präsident der europäischen Vereinigung für erneuerbare Energien. „2,7-mal mehr Solarstrahlung pro Quadratmeter als in Mitteleuropa. Und acht Millionen Quadratkilometer Wüstenfläche, obwohl 300.000 Quadratkilometer Solarfläche für den gesamten Weltbedarf ausreichen würde.“ [...]“ (Welt.de, 20.06.2009)

Der Bau von großen Wasserkraftanlagen und damit verbundenen Stauseen bedeutet einen schweren Eingriff in die Natur. Bisweilen werden dabei wertvolle natürliche Lebensräume zerstört. Der Betrieb von Wasserkraftwerken kann den Fischbestand im Gewässer gefährden, da die Turbinen eine Gefahr für die Tiere darstellen.

Position 3: Fossile Brennstoffe (Erdöl, Erdgas, Kohle)

In Deutschland lagern derzeit etwa 77 Milliarden Tonnen Braunkohle, von denen 53% (ca. 41 Milliarden Tonnen) mit dem Stand der heutigen Technik gewinnbar wären. Damit würden die Vorräte bei konstanter Förderung (2004: 181,9 Millionen Tonnen) noch für 225 Jahre ausreichen. Von den deutschen Steinkohlevorräten gelten rund 24 Milliarden Tonnen als gewinnbar. Angesichts einer aktuellen Förderquote von 25,7 Millionen Tonnen (2004) ergibt sich eine theoretische Reichweite von über 900 Jahren.

Ein einzelnes Steinkohlekraftwerk hat eine typische elektrische Leistung von bis zu 700 MW. Solch ein Kraftwerk benötigt etwa 1,8 Millionen Tonnen Kohle als Brennstoff im Jahr.

Das globale Ölfördermaximum wird auf das Jahr 2020 datiert. Nach dem Maximum wird mit sinkenden Fördermengen bei gleichzeitig steigendem Weltenergiebedarf gerechnet. Die statistische Reichweite (Reichweite bei derzeitigem Verbrauch und Preis) von Erdgas und Erdöl ist deutlich länger, aber ebenfalls so begrenzt, dass mittelfristig Alternativen notwendig sind.

Startart	**Anfahrzeiten***
Kaltstart nach weniger als 72h Stillstand	400 Minuten
Warmstart nach weniger als 48h Stillstand	280 Minuten
Heißstart nach weniger als 8h Stillstand	115 Minuten

**Die Anfahrzeit deckt die Dauer vom Anzünden des ersten Brenners bis zum Erreichen der Volllast ab.*

Tabelle E.10: Anfahrverhalten eines Kohlekraftwerks

Deutschland importierte im Jahr 2004 ca. 110,14 Millionen Tonnen Rohöl. Der wichtigste Erdöl- und Erdgas-lieferant für Deutschland ist mit etwa einem Drittel Russland, gefolgt von Norwegen mit etwa einem Viertel des Bedarfs. Der Anteil des aus deutschen Quellen gewonnenen Erdöls liegt bei etwa 3% des Verbrauchs.

Derzeit befinden sich 27 neue Kohlekraftwerke (überwiegend Steinkohle) in Deutschland im Bau bzw. in Planung.

Die Menge des in Lagerstätten enthaltenen Erdgases kann nur unsicher geschätzt werden. Die Schätzungen über die weltweiten Erdgasressourcen variieren zwischen 70.000-181.000 Milliarden Kubikmeter. Bei welt-weiter gleichbleibender Erdgasförderung entspricht dies einer statistischen Reichweite von etwa 62 Jahren.

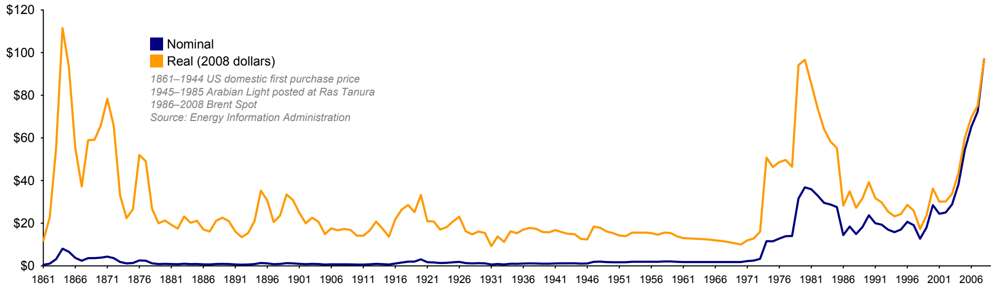


Abbildung E.11: Ölpreis nominal und inflationskorrigiert von 1861 bis 2007

Für Deutschlands Erdgasversorgung sind die wichtigsten Lieferländer: Russland (32%), Norwegen (26%) und die Niederlande (19%). Rund ein Fünftel des Bedarfs wird in Deutschland selber gefördert.

Zur Spitzendeckung, zum Ausgleich kurzfristiger Importstörungen und Bedarfsschwankungen werden in Deutschland ca. 18,6 Milliarden Kubikmeter Erdgas in Untergrundspeichern gelagert.

Der Wirkungsgrad von Kohlekraftwerken weltweit beträgt im Mittel 31%, in Deutschland liegt er bei 38%. Vor allem in Ländern wie China (durchschnittlicher Wirkungsgrad 23%) gibt es große Verbesserungspotenziale.

„[. . .] Als Lennart zum ersten Mal die Luft wegblieb, ergriff seine Eltern noch Panik. Sie packten den Säugling ins Auto und rasten zum Krankenhaus. Da wohnten sie gerade wieder sechs Monate in Lünen, Kreis Unna, am östlichen Rand des Ruhrgebiets. „Nun röcheln alle Kinder. Jeden Tag“, sagt Sara Köhler. [. . .] Familie Köhler hat den Feind ihrer Lungen längst ausgemacht: Es ist das Kohlekraftwerk vor ihrer Haustür, glaubt sie. [. . .] Seit Monaten gehen der 36-jährige Sozialarbeiter und die Pädagogik-Studentin zu Demonstrationen gegen neue Kohlekraftwerke, die älteren beiden Kinder kommen mit auf die Straße. „Wir tun, was wir können, aber es wird wohl nicht reichen.“

Denn der Kohlehunger der Energiekonzerne ist groß. Und Befürworter der Industrie sind zahlreich. Die Köhlers und ihre wachsende Bürgerinitiative argumentieren mit ihren Beobachtungen, sie erzählen Geschichten aus dem Kindergarten und der Schule. Eindeutige Zahlen können sie nicht liefern. Es existiert bundesweit keine Studie darüber, wie sich Kohlekraftwerke auf die Gesundheit der Anwohner auswirken können. [. . .]“ (Frankfurter Rundschau Online, 06.01.2009)

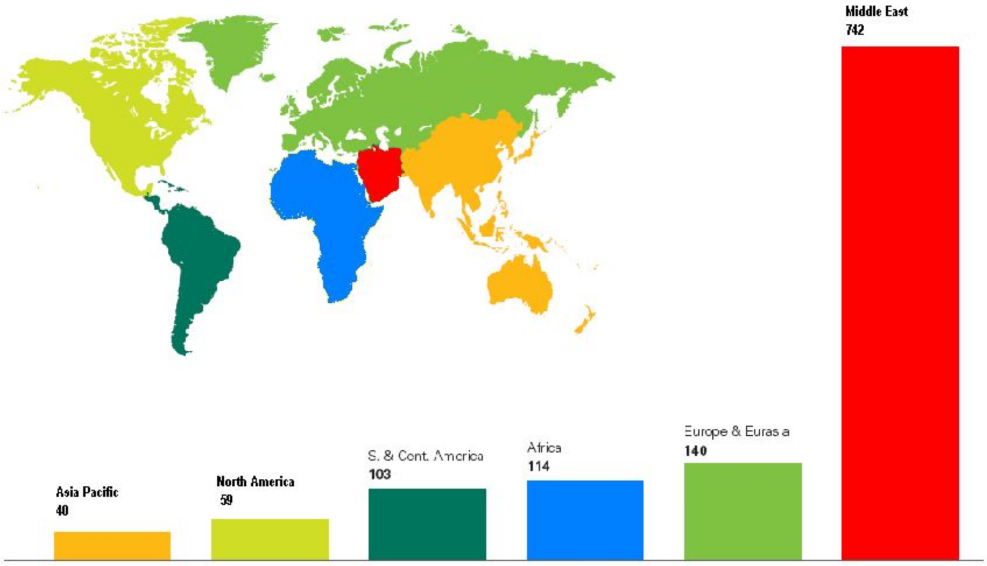


Abbildung E.12: Nachgewiesene Ölreserven in Millionen Barrel (by Pietros Sacanis)

F Materials of the Authoring Tool Test

F.1 LASAD Configuration Task 1

User definition

Please create the following three users:

Username	Password	Role
Bob	Bob123	Guest
Pia	456Pia	Guest
Tom	Tom654	Developer

Ontology definition

Please create a new ontology named "*Hypothetical Reasoning Ontology*".

The ontology should consist of three nodes and one relation. Each node should have a different border style.

The nodes should be titled as following:

1. *Hypothetical* (in red color, including a text field as well an internal link to a given text)
2. *Fact* (in orange color, including a text field)
3. *Test* (in blue color, including two labeled text fields (labels: “if” and “then”) as well as a drop down menu labeled “Confirmed?” with two options: Yes / No)

The following relation should be available:

1. *is connected to* (in green color, not directed and without any explanations, i.e. only the line without any additional panel that shows the title)

Template definition

Please define a template named “Hypothetical Reasoning Template”. The template should be based on the newly created ontology "Hypothetical Reasoning Ontologyänd allow only one user at a time. Furthermore, the following given text should be available:

Line No.	Content
1	Argumentation is essential in many aspects of life — in private as well as in business.
2	However, most people struggle to engage in reasonable argumentation. Thus it is important to support argumentation learning from early childhood on.
3	One way to teach argumentation is computer-based argumentation tools.

Session definition

Finally, the following pre-defined sessions based on the template “Hypothetical Reasoning Template” should be created:

1. Hypothetical Argumentation — Session 1 (Restricted to user “Bob”)
2. Hypothetical Argumentation — Session 2 (Restricted to user “Pia”)

F.2 LASAD Configuration Task 2

You plan to conduct a small lab study with three participants. In this lab study, the participants should argue about global warming. Therefore, the system should provide multiple workspace elements: First of all, they need to state hypotheses. These hypotheses must be supported by means of data. The data will be entered by the users as text and supported via web references to external material. In addition, each data should get a believability score ranging from 0 to 5 assigned. To connect the data to the hypotheses, there must be two kinds of relations possible: pro and con. In the con case, there should be a text field available that allows the user to clarify why it is con. All relation should have a direction. Further, nodes and relations should have a color on their own. Each node should have a different border to make them easier distinguishable.

The participants should use anonymous standard user accounts. These accounts should be named *lab1* to *lab3*. In addition, there should be another user account named *mod1* with additional rights for teachers. For all accounts, the passwords should be equal to the username.

All participants and the teacher should be able to work in the same argument session at the same time and there will be two runs. To allow for an easy coordination of activities they should be able to see who is active in the session. Additionally, the participants should be able to chat with each other in the system. Further, it should be clear who created which part of the argument graph und who is currently working on which part of the argument graph. The latter can be done via a tracking of the cursor movements.

F.3 Authoring tool questionnaire

1. I have been able to configure the LASAD framework by means of the authoring tool.

   Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree

2. I have been able to configure the LASAD framework by means of the authoring tool in an adequate time.

   Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree

3. I think the authoring tool of the LASAD framework is easy to use.

   Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree

4. Did you experience any problems? Can you describe these problems?

5. How do you think can the authoring tool be improved?
6. Do you have any additional comments?

G LASAD Configurations

G.1 Toulmin

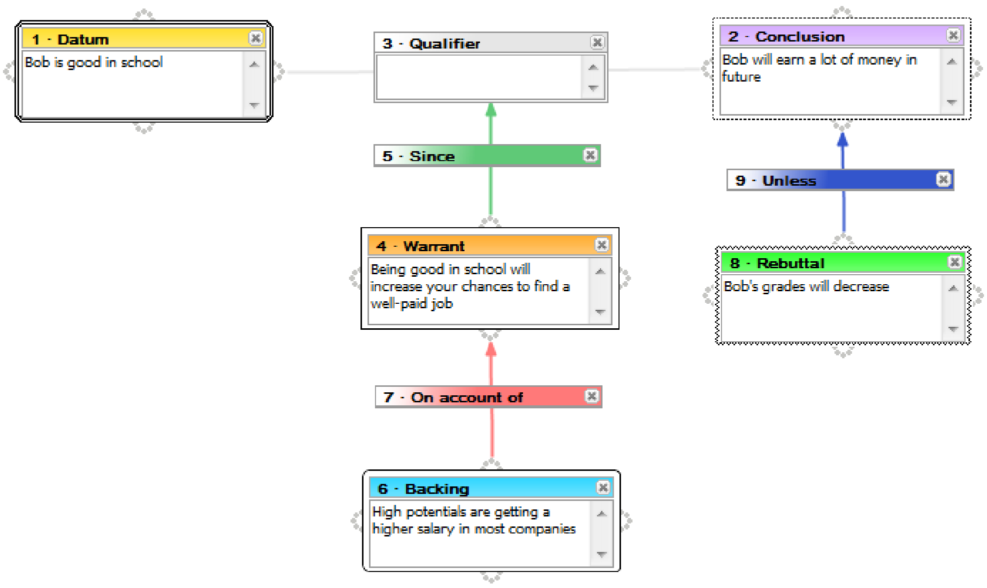


Abbildung G.13: Example of Toulmin configuration of LASAD

```
<ontology type="Toulmin (Single user)">

  <elements>

    <!-- Definition of nodes / boxes -->
    <!-- Each box has a text field -->
    <element elementtype="box" elementid="Backing" minquantity="0"
        maxquantity="0" quantity="0">
      <elementoptions heading="Backing"/>
      <uisettings height="200" width="200" background-color="#00CCFF"
          border="round"/>
      <childelements>
        <element elementtype="text" elementid="text" minquantity="1"
            maxquantity="1" quantity="1">
          <elementoptions texttype="textarea"/>
          <uisettings minheight="48" maxheight="160"/>
          <childelements/>
```

```
    </element>
  </childelements>
</element>

<element elementtype="box" elementid="Warrant" minquantity="0"
    maxquantity="0" quantity="0">
  <elementoptions heading="Warrant"/>
  <uisettings height="200" width="200" background-color="#FF9900"
      border="standard"/>
  <childelements>
    <element elementtype="text" elementid="text" minquantity="1"
        maxquantity="1" quantity="1">
      <elementoptions texttype="textarea"/><uisettings minheight="48"
          maxheight="160"/>
      <childelements/>
    </element>
  </childelements>
</element>

<element elementtype="box" elementid="Conclusion" minquantity="0"
    maxquantity="0" quantity="0">
  <elementoptions heading="Conclusion"/>
  <uisettings height="200" width="200" background-color="#CC99FF"
      border="dashed"/>
  <childelements>
    <element elementtype="text" elementid="text" minquantity="1"
        maxquantity="1" quantity="1">
      <elementoptions texttype="textarea"/><uisettings minheight="48"
          maxheight="160"/>
      <childelements/>
    </element>
  </childelements>
</element>

<element elementtype="box" elementid="Rebuttal" minquantity="0"
    maxquantity="0" quantity="0">
  <elementoptions heading="Rebuttal"/>
  <uisettings height="200" width="200" background-color="#00FF00"
      border="zigzag"/>
  <childelements>
    <element elementtype="text" elementid="text" minquantity="1"
        maxquantity="1" quantity="1">
    <elementoptions texttype="textarea"/>
    <uisettings minheight="48" maxheight="160"/>
    <childelements/>
    </element>
  </childelements>
</element>

<element elementtype="box" elementid="datum" minquantity="0"
    maxquantity="0" quantity="0">
  <elementoptions heading="Datum"/>
  <uisettings height="150" width="200" background-color="#FFD800"
      resizable="true" font-color="#000000" border="double"/>
  <childelements>
    <element elementtype="text" elementid="text" minquantity="1"
```

```
                maxquantity="1" quantity="1">
            <elementoptions texttype="textarea"/>
            <uisettings minheight="48" maxheight="160"/>
            <childelements/>
          </element>
        </childelements>
      </element>

      <!-- Definition of vertices / relations -->

      <!-- The qualifier relation contains may contain a text -->
      <element elementtype="relation" elementid="qualifier" minquantity="0"
           maxquantity="0" quantity="0">
        <elementoptions endings="false" heading="Qualifier"/>
        <uisettings linewidth="2px" linecolor="#E6E6E6" height="100" width=
             "165" background-color="#E6E6E6" resizable="false" font-color="
             #000000" border=""/>
        <childelements>
          <element elementtype="text" elementid="text" minquantity="1"
               maxquantity="1" quantity="1">
            <elementoptions texttype="textarea"/>
            <uisettings minheight="32" maxheight="32"/>
            <childelements/>
          </element>
        </childelements>
      </element>

      <!-- The other relations do not contain any child-elements, i.e. no
           text fields, etc. -->
      <element elementtype="relation" elementid="On account of" minquantity
           ="0" maxquantity="0" quantity="0">
        <elementoptions heading="On account of" endings="true"/>
        <uisettings width="160" height="200" background-color="#FF7979"
             linecolor="#FF7979" linewidth="2px"/>
        <childelements/>
      </element>

      <element elementtype="relation" elementid="Unless" minquantity="0"
           maxquantity="0" quantity="0">
        <elementoptions heading="Unless" endings="true"/>
        <uisettings width="160" height="200" background-color="#3355CC"
             linecolor="#3355CC" linewidth="2px"/>
        <childelements/>
      </element>

      <element elementtype="relation" elementid="Since" minquantity="0"
           maxquantity="0" quantity="0">
        <elementoptions heading="Since" endings="true"/>
        <uisettings width="160" height="200" background-color="#5FC977"
             linecolor="#5FC977" linewidth="2px"/>
        <childelements/>
      </element>

    </elements>

</ontology>
```

G.2 Belvedere

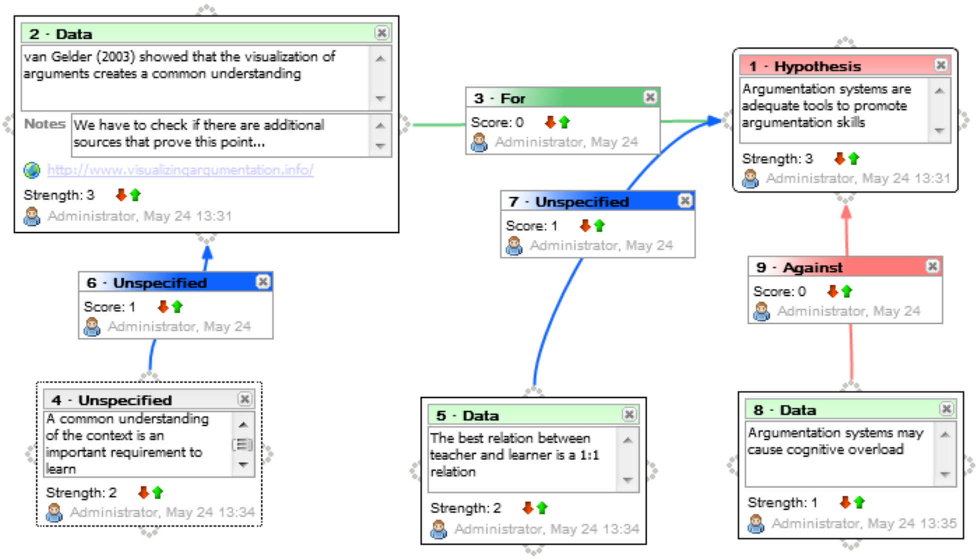


Abbildung G.14: Example of Belvedere configuration of LASAD

```
<ontology type="Belvedere">

  <elements>

    <element elementtype="box" elementid="data" minquantity="0"
        maxquantity="0" quantity="0">
      <elementoptions heading="Data"/>
      <uisettings width="180" height="250" resizable="true" border="
          standard" background-color="#c4ffc4" font-color="#000000"/>
      <childelements>
        <element elementtype="text" elementid="text" minquantity="1"
            maxquantity="1" quantity="1">
          <elementoptions texttype="textarea"/>
          <uisettings minheight="48" maxheight="96"/>
          <childelements/>
        </element>
        <element elementtype="text" elementid="notes" minquantity="0"
            maxquantity="1" quantity="0">
          <elementoptions texttype="textarea" label="Notes"/>
          <uisettings minheight="32" maxheight="32"/>
          <childelements/>
        </element>
        <element elementtype="url" elementid="url" minquantity="0"
            maxquantity="1" quantity="0">
          <elementoptions/>
          <uisettings minheight="16" maxheight="16"/>
          <childelements/>
        </element>
```

```
    <element elementtype="rating" elementid="strength" minquantity="1" maxquantity="1" quantity="1">
      <elementoptions score="0" minscore="-3" maxscore="3" label="Strength"/>
      <uisettings minheight="16" maxheight="16"/>
      <childelements/>
    </element>
    <element elementtype="awareness" elementid="author" minquantity="1" maxquantity="1" quantity="1">
      <elementoptions/>
      <uisettings minheight="16" maxheight="16"/>
      <childelements/>
    </element>
  </childelements>
</element>

<element elementtype="box" elementid="hypothesis" minquantity="0" maxquantity="0" quantity="0">
  <elementoptions heading="Hypothesis"/>
  <uisettings width="180" height="250" resizable="true" border="round" background-color="#FF8080" font-color="#000000"/>
  <childelements>
    <element elementtype="text" elementid="text" minquantity="1" maxquantity="1" quantity="1">
      <elementoptions texttype="textarea"/>
      <uisettings minheight="48" maxheight="96"/>
      <childelements/>
    </element>
    <element elementtype="text" elementid="notes" minquantity="0" maxquantity="1" quantity="0">
      <elementoptions texttype="textarea" label="Notes"/>
      <uisettings minheight="32" maxheight="32"/>
      <childelements/>
    </element>
    <element elementtype="url" elementid="url" minquantity="0" maxquantity="1" quantity="0">
      <elementoptions/>
      <uisettings background-color="#FFFFFF" font-color="#000000" minheight="15" maxheight="15"/>
      <childelements/>
    </element>
    <element elementtype="rating" elementid="strength" minquantity="0" maxquantity="1" quantity="1">
      <elementoptions label="Strength" score="0" minscore="-10" maxscore="10"/>
      <uisettings background-color="#FFFFFF" font-color="#000000" minheight="15" maxheight="15"/>
      <childelements/>
    </element>
    <element elementtype="awareness" elementid="awareness" minquantity="1" maxquantity="1" quantity="1">
      <elementoptions text=""/>
      <uisettings background-color="#FFFFFF" font-color="#000000" minheight="15" maxheight="15"/>
      <childelements/>
    </element>
```

```
  </childelements>
</element>

<element elementtype="box" elementid="unspecified" minquantity="0"
    maxquantity="0" quantity="0">
  <elementoptions heading="Unspecified"/>
  <uisettings width="180" height="250" resizable="true" border="
      dashed" background-color="#E6E6E6" font-color="#000000"/>
  <childelements>
    <element elementtype="text" elementid="text" minquantity="1"
        maxquantity="1" quantity="1">
      <elementoptions texttype="textarea"/>
      <uisettings minheight="48" maxheight="96"/>
      <childelements/>
    </element>
    <element elementtype="text" elementid="notes" minquantity="0"
        maxquantity="1" quantity="0">
      <elementoptions texttype="textarea" label="Notes"/>
      <uisettings minheight="32" maxheight="32"/>
      <childelements/>
    </element>
    <element elementtype="url" elementid="url" minquantity="0"
        maxquantity="1" quantity="0">
      <elementoptions/>
      <uisettings background-color="#FFFFFF" font-color="#000000"
          minheight="15" maxheight="15"/>
      <childelements/>
    </element>
    <element elementtype="rating" elementid="strength" minquantity="0
        " maxquantity="1" quantity="1">
      <elementoptions label="Strength" score="0" minscore="-10"
          maxscore="10"/>
      <uisettings background-color="#FFFFFF" font-color="#000000"
          minheight="15" maxheight="15"/>
      <childelements/>
    </element>
    <element elementtype="awareness" elementid="awareness"
        minquantity="1" maxquantity="1" quantity="1">
      <elementoptions text=""/>
      <uisettings background-color="#FFFFFF" font-color="#000000"
          minheight="15" maxheight="15"/>
      <childelements/>
    </element>
  </childelements>
</element>

<element elementtype="relation" elementid="for" minquantity="0"
    maxquantity="0" quantity="0">
  <elementoptions heading="For" endings="true"/>
  <uisettings width="140" height="120" resizable="false" border=""
      background-color="#5FC977" font-color="#000000" linewidth="2px"
       linecolor="#5FC977"/>
  <childelements>
    <element elementtype="rating" elementid="strength" minquantity="1
        " maxquantity="1" quantity="1">
      <elementoptions score="0" minscore="-10" maxscore="10"/>
```

```
        <uisettings background-color="#FFFFFF" font-color="#000000"
            minheight="15" maxheight="15"/>
        <childelements/>
      </element>
      <element elementtype="awareness" elementid="awareness"
          minquantity="1" maxquantity="1" quantity="1">
        <elementoptions text=""/>
        <uisettings background-color="#FFFFFF" font-color="#000000"
            minheight="15" maxheight="15"/>
        <childelements/>
      </element>
    </childelements>
  </element>

  <element elementtype="relation" elementid="against" minquantity="0"
      maxquantity="0" quantity="0">
    <elementoptions heading="Against" endings="true"/>
    <uisettings width="140" height="120" resizable="false" border=""
        background-color="#FF7979" font-color="#000000" linewidth="2px"
         linecolor="#FF7979"/>
    <childelements>
      <element elementtype="rating" elementid="strength" minquantity="1
          " maxquantity="1" quantity="1">
        <elementoptions score="0" minscore="-10" maxscore="10"/>
        <uisettings background-color="#FFFFFF" font-color="#000000"
            minheight="15" maxheight="15"/>
        <childelements/>
      </element>
      <element elementtype="awareness" elementid="awareness"
          minquantity="1" maxquantity="1" quantity="1">
        <elementoptions text=""/>
        <uisettings background-color="#FFFFFF" font-color="#000000"
            minheight="15" maxheight="15"/>
        <childelements/>
      </element>
    </childelements>
  </element>

  <element elementtype="relation" elementid="unspecified" minquantity="
      0" maxquantity="0" quantity="0">
    <elementoptions heading="Unspecified" endings="true"/>
    <uisettings width="140" height="120" resizable="false" border=""
        background-color="#0E63FF" font-color="#000000" linewidth="2px"
         linecolor="#0E63FF"/>
    <childelements>
      <element elementtype="rating" elementid="strength" minquantity="1
          " maxquantity="1" quantity="1">
        <elementoptions score="0" minscore="-10" maxscore="10"/>
        <uisettings background-color="#FFFFFF" font-color="#000000"
            minheight="15" maxheight="15"/>
        <childelements/>
      </element>
      <element elementtype="awareness" elementid="awareness"
          minquantity="1" maxquantity="1" quantity="1">
        <elementoptions text=""/>
        <uisettings background-color="#FFFFFF" font-color="#000000"
```

```
              minheight="15"  maxheight="15"/>
          <childelements/>
        </element>
      </childelements>
    </element>

  </elements>

</ontology>
```