# IMPROVING PEER LEARNING AND KNOWLEDGE SHARING IN STEM COURSES VIA PATTERN BASED GRAPH VISUALIZATION

by

**Emilio Francisco Zegarra**

B.S. Industrial Engineering, Purdue University, 1999

M.S. Industrial Engineering, University of Pittsburgh, 2002

M.S. Information Technology – Software Engineering, Carnegie Mellon University, 2004

Submitted to the Graduate Faculty of

The Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

**Doctor of Philosophy**

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

THE KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Emilio Francisco Zegarra

It was defended on

July 10, 2015

and approved by

Shi-Kuo Chang, Department of Computer Science

Jingtao Wang, Department of Computer Science

Diane Litman, Department of Computer Science

Christian Schunn, Department of Psychology

Jennifer Iriti, Learning Research and Development Center

Dissertation Advisor: Shi-Kuo Chang, Department of Computer Science

**IMPROVING PEER LEARNING AND KNOWLEDGE SHARING IN STEM**

**COURSES VIA PATTERN BASED GRAPH VISUALIZATION**

Emilio Francisco Zegarra, PhD

University of Pittsburgh, 2015

High quality education in Science, Technology, Engineering and Math (STEM) majors expects not only the acquisition of comprehensive domain knowledge, but also the mastery of skills to solve open-ended and even ill-defined problems in real world. Problem-based Learning (PBL) is usually adopted to achieve such goals by encouraging students to learn by solving real-life problems. However, successful PBL requires sustained and in-depth involvement of faculty members, hence making PBL not scalable. Even though discussion forums and Q&A systems can help address the scalability problem of faculty involvement on large class sizes, it introduces new problems. First, as knowledge bases grow in size, the sheer size of the accumulated knowledge makes it harder to locate the desired information. Second, existing knowledge discovery techniques do not provide effective facilities for the capture and reuse of solutions to recurring problems.

To address these challenges, we developed MicroBrowser, an innovative and interactive Question & Answer (Q&A) system augmented with pattern-based expertise-sharing interfaces and 2D knowledge graph discussion visualization. MicroBrowser provides a set of pattern-based expertise-sharing interfaces to allow both learners and instructors to refine, reuse, and share knowledge. MicroBrowser also allows learners to browse and navigate important discussions based on topic similarity encoded by node proximity in a knowledge graph.

Results of empirical evaluations of our proposed solution show that ask difficulty improves with MicroBrowser when compared with a state-of-the-art Q&A system for knowledge discovery and reuse tasks. In addition, success rate for knowledge discovery tasks using keywords was higher with MicroBrowser. Moreover, we show that, students found the pattern-based expertise-sharing interface easy to use and were able to contribute new knowledge in the form of new knowledge connections and even recommend new design patterns.

**TABLE OF CONTENTS**

# LIST OF TABLES

# LIST OF FIGURES

**PREFACE**

I would like to thank my committee advisor Dr. Chang and my committee members Dr. Wang, Dr. Litman, Dr. Schunn, and Dr. Iriti for their support and guidance as I worked towards completing the requirements for the degree. I also would like to thank all the study participants for their commitment to validate the feasibility of my work.

I would like to dedicate this thesis first and foremost to my wife Dr. Bedda Lynn Rosario who provided me the support, not only emotional but with everything else anyone would need to cope with this degree. She even helped me with reviewing my statistical analysis. It has been a long road and I am very happy to be close to the end. To my daughter Ana Bedda who always reminds me how to continue living a happy life. To my son Javier Emilio, who at only 3 months of age is already wondering when will I be done and of course, when will we feed him next. A big THANK YOU to all three of you.

I would also like to thank my parents (Santiago and Elena), brothers (Santiago and Miguel) and sisters (Elena and Cristina) for cheering for me. To my family in Puerto Rico and in Peru, friends and colleagues who have provided support by reviewing my thesis, approving my expenses, listening to my complaints, or just wondering when I will be done.

# 1.0    INTRODUCTION

Work on improving the education of students in all the fields of Science, Technology, Engineering and Math (STEM) majors has allowed students to acquire important skills in such fields.  Yet, upon graduation, students in those fields are presented with open-ended and ill-defined problems for which they might not be prepared.  Innovative techniques such as Problem-based Learning (PBL) aim to address these limitations by encouraging students to learn by addressing everyday problems (Hung, Jonassen, & Liu, 2012). Examples of the use of PBL techniques applied to fields of STEM education are reflected in Capstone Projects.  Capstone Projects stimulate the development of communication, problem solving and teamwork skills among students as part of their educational curriculum (Franchetti & Hefzy, 2012)(Dutson, Todd, Magleby, & Sorensen, 1997).    However, Capstone Projects require the in-depth participation of faculty advisors (Massie & Massie, 2006). Faculty advisors cannot expect students to perform on their own without providing any guidance.

While PBL techniques are feasible for small faculty-to-student ratios, as the ratios become larger, faculty involvement becomes more difficult resulting in a longer feedback loop. Peer learning techniques such as Peer Instruction can be an alternative to faculty involvement by allowing students to participate in their active learning (Crouch & Mazur, 2001)(C. B. Lee, Garcia, & Porter, 2013).  However, for peer learning to be effective, it requires the availability of supporting infrastructure that facilitates the exchange of knowledge between peers.  To address

this challenge, discussion forums and Question & Answer (Q&A) systems (Coetzee, Fox, & Hearst, 2014)(Huang, Dasgupta, Ghosh, Manning, & Sanders, 2014) have been used in the context of online education via Massive Open Online Courses (MOOCs). These systems have been found to be a common source of knowledge for students when completing homework tasks (Sonnino, 2013)(Seaton, Bergner, Chuang, Mitros, & Pritchard, 2014)(Huang et al., 2014), and for interaction with instructors (Huang et al., 2014). Using such systems, students are able to receive feedback faster as the knowledge exchange does not only occur between the student and the faculty, but it also enables peer-to-peer learning.

Even though discussion forums and Q&A systems can help address the scalability problem of faculty involvement on large class sizes by facilitating peer learning and knowledge discovery, it introduces new problems. First, as knowledge bases grow in size, the sheer size of the accumulated knowledge makes it harder to locate the desired information. Second, existing knowledge discovery techniques do not provide effective facilities for the capture and reuse of solutions to recurring problems under different contexts.

The success of an online knowledge base depends on student participation. One approach to bootstrap students' participation is to demonstrate value and encourage usage. This is done by making it easy for students to interact with the online knowledge base and by facilitating the ability to discover desired information. Q&A systems allow students to post their questions and from the answers received, the student can select the most appropriate one. Current state-of-the-art Q&A systems such as StackOverflow[1] and Yahoo Answers![2] use text-based browsing schemes for navigating and discovering questions and answers. The simplicity of text-based browsing schemes makes the interaction process rather easy. Yet, by simplifying

---

[1] http://www.stackoverflow.com
[2] http://answers.yahoo.com

the access method, we narrow the view to textual representations of the search results and fail to convey additional information that might be relevant. The use of visualization techniques has been applied to discussion forums to improve text-based browsing schemes. ForAvis (Wanner, Ramm, & Keim, 2011) used visualization to analyze forum activity using a combination of sequential order, color mapping and faceted browsing, iBlogVis (Indratmo & Gutwin, 2008) looked at blog entries and used different shapes to denote entries, comments and number of words, VIDI Toolbar (Trampuš & Grobelnik, 2010) and Fortuna et al. (Fortuna, Grobelnik, & Mladenić, 2005) used visualization to explore topic models extracted from text documents such as discussion threads, and Anagora (Giguet & Lucas, 2009) visualized discussion forums by their chronological properties. The visualization techniques adopted by these prior works such as color coding, shape encoding, and topic modeling provided us with inspiration to facilitate knowledge discovery in large knowledge bases.

Techniques such as Frequently Asked Question (FAQ), pattern theory (Alexander, 1999) and the use of design patterns (Durdik & Reussner, 2013) are techniques used for knowledge reuse. FAQ lists provide effective means for capturing answers to commonly asked questions. While they facilitate knowledge reuse, their simple structures capture limited information and their specific contexts limit their reusability. Christopher Alexander (Alexander, 1999) described design patters as a shared language used to communicate proven solutions to recurring problems. This ability to help identify solutions to recurring problems reusable form has facilitated their adoption in other domains. For example, User interface design patterns have also been created and provide best practices for the implementation of graphical user interfaces (Vanderdonckt, 2010)(Lin & Landay, 2008). Design patterns have also been incorporated into educational settings. Astrachan et al. developed instruction material and frameworks that

incorporate design patterns to facilitate teaching object oriented programming to Computer Science students (Astrachan, Mitchener, Berry, & Cox, 1998). Muller (Muller, 2005) proposed Pattern Oriented Instruction as a pedagogical technique that recommends adopting design pattern in course instruction with a goal of improving problem-solving skills in students. More recent work has focused on adopting design patterns in teaching parallel and distributed programming to undergraduate students (Brown, Adams, Ferner, Shoop, & Wilkinson, 2014). Pedagogical Patterns were also defined to assist instructors in preparing instructional material based on learned experience from expert instructors (Carle, Clancy, & Canny, 2007) (Köppe & Utrecht, 2011).

As an example of how we can integrate design patterns into a Q&A system, let us assume that a new question posted by a student is similar to another question already answered in the system except that some parameters are different. Let us also assume that the existing question is already associated to a design pattern. We can then associate the new question to the same design pattern by assuming that the question is the recurring problem, the solution is the approach to answering the question and the context is the education setting. By using design patterns we can represent the question and its answers in an abstract form to allow us to generalize its solution to different problems. Likewise, if a student notices an unanswered question that matches a design pattern, the student can refer to the design pattern solution as well as instances of the design pattern in the form of other questions, to compose an appropriate answer to the question. Therefore, the concept of design patterns provides an innovative way to bootstrap students in the process of knowledge generation and reuse.

Even though prior work has explored the integration of design patterns to scaffold the learning process and the use of visualization techniques to facilitate knowledge discovery in

discussion forums, no prior work has combined them to take advantage of their benefits of knowledge reuse and discovery. Therefore, in this work we have developed an innovative pattern-based visualization system, called MicroBrowser, to improve knowledge discovery and reuse by integrating design patterns and using 2D visualization techniques to facilitate knowledge discovery. The major contribution of this work is the integration of design patterns with an interactive knowledge graph to facilitate knowledge reuse and knowledge discovery in large knowledge bases.

## 1.1    OVERVIEW OF CONTRIBUTIONS

The main contribution of this thesis is an innovative approach that integrates design patterns with an interactive knowledge graph to facilitate knowledge reuse and knowledge discovery in large knowledge bases. In this work we designed and implemented a streamlined knowledge-graph navigation system augmented with 2D discussion visualizations and design pattern based expertise-sharing interfaces in fields of STEM majors called MicroBrowser. Inspired by the knowledge generation capabilities in Q&A systems and the reusability features of FAQ lists, we incorporate design patterns as reusable knowledge nuggets that provide solutions to recurring problems. Results from a user study and pilot study suggest that by bootstrapping the knowledge discovery and generation process via design patterns, students can effectively complete educational tasks that were difficult to complete using traditional Q&A systems. We also propose to improve knowledge discovery in large knowledge bases by the use of visualization techniques in combination with text-based browsing schemes and knowledge timeline. We used visualization techniques to take advantage of humans' visual exploration capabilities and

5

demonstrate that students can effectively discover knowledge from discussion threads in the form of content, attributes, relationships, and time information. To evaluate our system, we designed and developed an innovative streamlined knowledge-graph navigation system called MicroBrowser augmented with 2D discussion visualizations and pattern-based expertise-sharing interfaces. MicroBrowser allows learners to browse and navigate important discussions based on topic similarity encoded by node proximity in a knowledge graph. MicroBrowser also provides a set of pattern-based expertise-sharing interfaces to allow both learners and instructors to refine, highlight, reuse, and share major findings in PBL.

Finally, in a user study performed with students in Computer Science and Information Science fields, we demonstrate that our proposed system improves the success rate and lowers difficulty for completing knowledge discovery and exploration tasks when compared to a state-of-the art Q&A system. We also demonstrate that students in Computer Science and Information Technology fields effectively integrate the use of design patterns for knowledge discovery, generation and reuse.

## 1.2    THESIS OUTLINE

The remainder of the dissertation is organized as follows: Chapter 2 provides background information on online knowledge sharing communities and presents a taxonomy we defined for classifying them. We also describe discussion forums and Q&A systems. Chapter 3 describes an exploratory evaluation on the use of online knowledge communities in an education setting to understand motivators and inhibitors to participation. In Chapter 4, we address the scalability problem, describe our interactive knowledge visualization system and introduce the integration

of design patterns for knowledge generation and reuse. In Chapter 5, we present evaluation results on how MicroBrowser promotes peer learning and knowledge discovery and compare it to a traditional Q&A system. In Chapter 6, we describe results of a pilot study whose aim was to assess how student-generated design patterns enhance peer-learning. Finally, Chapter 7 summarizes the major discoveries that we obtained from our work and suggests future research that could extend the work of this dissertation.

## 2.0    BACKGROUND

The work at hand is placed in the context of literature related to online knowledge sharing communities, technologies and approaches to support forum visualization and the use and integration of design patterns.  The following subsections describe the areas in details.

## 2.1    KNOWLEDGE SHARING COMMUNITIES

When users need help or have a question, they tend to reach out to a trusted source to obtain answers or advice (Girgensohn & Lee, 2002).  The power of social collaboration enables users to contribute knowledge for the greater good of the community.  In addition, the growth of the internet and the availability of search crawlers have given students the tools to search for content themselves.  By doing this, students have the content but there is no easy way to trust its source. Additionally, when students have a question, they need to decompose the question into searchable keywords that can be submitted to a query engine, review the returned results and decide on the best answer by basically learning about the entire context of the search result (Agichtein, 2008).  For example, to find out "who was the first president?" the user first needs to decompose the question into keywords for identifying the question. After performing the search, the user needs to determine which search result is relevant. Then, the user has to possibly read about the entire history of the country as well as who were the other presidents.  While this

approach has worked well for now, it comes with an increase in the mental load a user needs to acquire before finding the correct answer to his question. Techniques such as Slow Intelligence Systems have used ontologies as a means of knowledge representation and ontology filtering as a means to finding desired information. However, their main advantage is also their limitation since knowledge needs to be represented as ontologies which differ from the unstructured syntax of the Web (Chang, Zegarra, et al., 2010)(Chang, Colace, Santo, Zegarra, & Qie, 2010)(Zegarra, Colace, De Santo, & Chang, 2010).

The increased use of the Web as a knowledge repository and social medium has led to the proliferation of online Knowledge Sharing Communities (KSC) such as Discussion Forums and Question and Answer (Q&A) systems. These communities rely on its members to create the knowledge by following a question and answer format with the knowledge created becoming available to all members of the community(Dror, Koren, Maarek, & Szpektor, 2011). Following a Question and Answer format becomes more intuitive to users. When users have a need to know something about a topic, it is easier to represent their inquiries in the form of a question and ask them to a trusted source rather than using a search engine and filtering on the results.

For this reason, students have been turning to online Knowledge Sharing Communities such as Yahoo! Answers("Yahoo! Answers," n.d.), WikiAnswers, Quora(Wang, Gill, & Mohanlal, 2013), and AskVille ("AskVille," n.d.), to obtain answers to questions on diverse topics (Agichtein, 2008). Knowledge Sharing Communities aim to become repositories of knowledge with a familiar Q&A format. They exploit the power of human knowledge to satisfy a broad range of users' information needs by handling factual as well as complex questions that could be very difficult, if not impossible, to answer using conventional web search engines (Pera, 2011).

### 2.1.1   Use of Computer Mediated Communication in Educational Settings

There has been a growing reliance on Computer Mediated Communication (CMC) systems and in particular asynchronous systems such as discussion forums and Q&A systems, to support learning and communications in traditional face-to-face and online classes.  In a study performed among Computer Science and Software Engineering students about the use of discussion forums in classroom settings found that students felt that discussion forum were useful as a source of help (Sheard, 2004).    In a traditional classroom, if there is even an opportunity for discussion, students must react quickly if they are to contribute. The asynchronous environment provided by Q&A systems, however, affords students the time for thoughtful analysis, reflection, and composition as their discussion of an issue evolves.  There are many differences in the use of Q&A systems in public versus educational domains.  One such difference in using Q&A systems in educational domains relies on the fact that its members have the common property of belonging to the same class or workgroup, forming an Educational Knowledge Sharing Community (eKSC).  This common bond, however, is finite as it is tied to either a semester or project lifetime, after which time, the bond disappears.  An eKSC is one that is created in order to support a particular course at an educational institution.  As such, issues such as short engagement time and focused questions are important.  Concerning short engagement time, a course tends to last for example a semester, quarter or possibly up to a year.  After the course ends, the members of the knowledge community would withdraw their participation along with the completion of the course.  Therefore, issues such as motivation and early bootstrapping become important, as you need to ensure that members are engaged from the start in order to achieve the most benefit of the community.  Since the knowledge community is associated to a particular course, the knowledge exchanged is relevant to that course material.  The content

created thus would be very specialized and focused on the students only. Community members would need the support of experts, such as the instructor and Teaching Assistants, and of their fellow classmates to build and share the knowledge base.

## 2.2    DISCUSSON FORUMS AND Q&A SYSTEMS

Several techniques that leverage technology have been implemented to support online KSCs. Two such techniques that are in common use are Discussion Forums and Q&A systems. We now describe each one of the systems and how they have been used in educational settings.

### 2.2.1    Discussion Forums

Discussion forums allow users to exchange (read and post) messages asynchronously. Discussion forums (sometimes referred to simply as Forums) are like Usenet "newsgroups," the original Internet discussion groups. Forums enable the exchange of knowledge between forum members on a variety of topics by posting questions and allowing other users to send responses. Most forums categorize postings by special interests. Generally, posting to newsgroups are not moderated. The main differences between newsgroups and online discussion forums are on how you access them as well as how they are moderated.  Online discussion forums are available only online and users access them via a web browser. On the other hand, newsgroups are public systems that are accessed online or through a newsgroups reader application. Anybody with access to a system that can connect to the newsgroups can post a question. In contrast, online discussion forums require users to be members of the Web site before they can submit a post.

The use of discussion forums and newsgroups to support and enhance learning has been explored since the early 1990s (Chizmar & Williams, 1996). Most recently, Massive Open Online Course (MOOC) systems such as Coursera and EdX, host hundreds to thousands of students in their courses. At such a scale, face-to-face interaction is nearly impossible thus making the use of discussion forums when interacting with students essential (Coetzee et al., 2014). Forums have thus become a critical component in MOOCs and the primary means of interaction between instructors and students (Huang et al., 2014).

### 2.2.2 Question and Answer Systems

Question and Answer (Q&A) systems allow users to seek and browse knowledge in the form of a natural language question. All Q&A sites have an interface designed for asking and answering questions (Harper, Raban, Rafaeli, & Konstan, 2008). Participants of Q&A sites, such as knowledge providers, contribute their answer to the community in the form of a self-contained answer (Agichtein, 2008). Questions and answers made available on a Q&A site become the knowledge of the community. Thus, as the number of questions with answers increases, the amount of knowledge shared among all the members in the community grows. Q&A sites have been able to increase their reach by making their capabilities available online.

One example of the use of Q&A systems in educational settings is the web-based tool Piazza[3]. Piazza was designed for educators to integrate into their courses; it allows participants to ask and answer questions. Different from traditional Q&A systems, Piazza allows students to ask questions but they can receive only two answers: one from other students and another from

---

[3] http://www.piazza.com

the instructor. If other students want to submit an answer, they must edit the student answer using a wiki-style editor. Aritajati and Narayanan (Aritajati & Narayanan, 2013) proposed Green Dolphin (GD) as an alternative Q&A system to support the collaborative learning of programming. In GD, students can ask other students or ask a system-identified pool of "experts".

## 2.3    DISCUSSION FORUM AND Q&A SYSTEMS CLASSIFICATION

In Section 2.1, we described KSCs; Section 2.2 focused on discussion forums and Q&A systems. The rapid proliferation and adoption of discussion forums and Q&A systems has made it hard for researchers to classify them so that similarities can be identified. Therefore, in this section we try to address this by first identifying the main characteristics that these systems can have and then grouping these characteristics into dimensions, as shown in Table 1. Using these defined dimensions, we aim to classify different implementations of discussion forums and Q&A systems to gain an understanding of their design space. As expected, because of the large similarity among them, it is feasible that the systems could adopt several dimensions as part of their design. We identified the following dimensions: Answerer selection, Member Reputation, Motivational, Communication, Question Type, Moderation, Best Answer Selection and Membership Scope.

### 2.3.1 Answerer Selection Dimension

Once the KSC receives a question post, it must determine who should answer it. Making the determination of what member should be answering has a profound effect on the quality as well as accuracy of the answers.

- Expert - These KSC rely on experts to provide answers to the questions. In this model, members submit questions to a category and the system selects an expert based on the selected category or their personal preferences. The expert might be an employee of the community system (but not necessarily). However, the system will only allow experts to post answers to questions. The expert researches the question for the users and provides the answer. One such example is Google Answers ("Google Answers," n.d.) which provides this type of service, with the further restriction that it only supports one answer per question, which comes from the expert assigned.

- Digital Reference Service – This type of KSC is very similar to the expert one, however, it relies on a structured process to handle questions asked by community members. These KSCs normally provide library services with a librarian handling the question posted by the members (Pomerantz, Nicholson, Belanger, & David Lankes, 2004). Similar to an expert system, only the researcher assigned to the question might submit an answer to it. Because of the limited number of resources available to research a question, there are lower numbers of answers. In addition, the categories available are limited to the expertise of the librarians. However, the member that asked the question expects to receive better and more accurate answers given the research time devoted to the question.

- Community Model – This is the most common implementation of Q&A sites. They typically leverage the power of all its members to research and provide answers to a question. It is

typical for this type of model to receive multiple answers. For this reason, some KSCs might implement different approaches to identify the best answer.

- Member Matching – In the member matching approach, the KSC filters the questions and presents them only to those members that can be related to it. One example of this approach is the Multi-Channel Recommender (MCR) system (Dror et al., 2011) which uses question and user preference attributes available to identify a list of potential answerers. Once the system identifies the prospective answerers list, it makes the question visible to those members only. One of the advantages of this approach is that users would only see unanswered questions that the system believes they can answer or relate to. The disadvantage of this approach is that it reduces the number of possible answerers for a question, as not every member would be notified of the question posted.

- Question Matching – In this approach, the KSC uses search algorithms in order to find past questions that are similar to the posted question and returns their answers. If answers are found, they are presented as answers to the posted question. An example of this approach is the question answering refinement method called QAR proposed by Pera and Ng (Pera, 2011). The QAR method first finds past answers to archived questions posted on Yahoo! Answers similar to the posted question. After the QAR method finds a list of questions, it uses an answer-ranking algorithm to select the top 10 answers and presents them as answers to the question posted. A problem with this method is that the answers that the system may find could be irrelevant and/or poorly written even if their associated questions are relevant (Suryanto, Lim, Sun, & Chiang, 2009).

### 2.3.2    Member Reputation Dimension

The reputation dimension provides a way for the KSC to build trust among the community members.  Its main goal is to provide a way for members of the community to know whom to trust (Dellarocas, 2010). KSCs employ different means to achieve this trust. The assumption is that trusted sources are likely to provide better answers (Fichman, 2011).    The following dimensions are samples of reputation metrics.

- Badges – Badges are similar in concept to obtaining medals.  In order to obtain a badge, the member must demonstrate a level of activity that is above the expected from a traditional user, thus representing interest of that activity.  Badges are usually shown as graphical icons indicating completion of an activity or achievement level.  Different from traditional game levels, members can acquire them in any order.  By using badges, other members can make inferences about the accomplishments of a particular member without having to actually read the member's profile or review his activity levels.

- Ranking on leaderboard – With this approach, members compare themselves against other members based on the number of achievement points accumulated.  One such example is Yahoo! Answers (Dror et al., 2011) which provides such a ranking not only to motivate users but also to build trust for its top contributors.  By ranking higher on the leaderboard, a member shows his commitments and knowledge contributions, and other members have a means to recognize it as they would be able to associate the difficulty of showing up on the ranks.  Leaderboards are public to all members and are typically displayed.

- Profile – A profile provides information about a user or group.  KSCs identify users by their handle or username.  The username or handle can be as simple as an email address, an alias, or a random identifier.  Because a handle or username might not provide enough information

about a user, a profile provides additional information to other members of the KSC (Girgensohn & Lee, 2002). Profiles can sometimes include a picture or avatar giving other community members a more personalized interaction.

- Total points – The use of points to track activity is one of the easiest ways to introduce reputation on a KSC. In essence, a community specifies the number of points a member would either gain or lose after performing an activity. The KSC displays these points next to the username or handle whenever the user contributes to the community. In this way, other members can measure how active a member is.

- External authority or credibility – In some cases, the members know each other outside of the context of the community either personally or by reference. In such scenario, the members of the community rely on this pre-existing trust to determine when the member should be trusted. However, to build this trust, the member must be identifiable with its external personality. This usually works well in small KSCs where members know each other or in specialty KSCs where known experts exist.

- Number of best answers – This metric represents how many members agree that a particular member has provided best answers. This is a very difficult metric to obtain as the number of members increases and the possibility of receiving multiple answers to a question increases.

### 2.3.3 Motivational Dimension

Simply because a community member posted a question to a KSC does not imply that a response would be received (Dearman & Truong, 2010). For this reason, KSCs include motivational features or incentives into their design to ensure their success (Dearman & Truong, 2010) (Festinger, 1954). Research literature suggests that the majority of the content is generated by

17

few users, sometimes referred to as the power law distribution of participation (Lampe, Wash, Velasquez, & Ozkaya, 2010). Additionally, a term called "social loafing" refers to the fact that some members expect other members of a group to contribute. What KSCs want to prevent is hesitation to contribute out of fear of criticism, or of misleading the community members by not being sure that their contributions are important, completely accurate, or relevant to a specific discussion (Ardichvili, Page, & Wentling, 2003). Below we describe several motivational dimensions.

- Member Altruism - The literature on KSCs has identified that one of the reasons community members participate is because of a sense of wanting to help others (Ardichvili et al., 2003)(Lampe et al., 2010). This help translates into an altruistic desire of experts to share their knowledge (Ardichvili et al., 2003)(Nam, Arbor, & Ackerman, 2009). In such a manner, KSCs benefit from this altruism by allowing users to freely post questions and answers.

- Economic Incentives – Economic incentives refer to the notion of being financially rewarded or punished for using the system. Just like there are some members willing to help others for no gain, some members want to be compensated for their time and effort. Users perceive a benefit when asking a question. In such a manner, they are willing to pay to ask a question (Hsieh & Counts, 2009)(Hsieh, Kraut, & Hudson, 2010). In (Hsieh et al., 2010), the authors found that users are more likely to pay to post a question when the type of question was factual or asked for advice. They also found that there is a correlation between the amount of the payment and the perceived difficulty of the question. Answering a question requires time and resource commitments from the potential experts. In such a way, members are willing to provide a better answer if there is going to be some kind of reward associated with it. In

general, paid questions had more answers than free questions (Hsieh et al., 2010)(Hsieh & Counts, 2009)(Harper et al., 2008).

- Gaming – Members want to enjoy using a KSC. Prior research demonstrates that member engagement increases once gaming concepts are introduced (Farzan, DiMicco, & Millen, 2008). One of the important characteristics in the use of game design is that you want to motivate good behavior and discourage bad behavior. In that sense, gaming elements need to be included such that participants are rewarded when they perform actions that contribute to knowledge contribution while penalties, such as losing points, are applied when undesired activities are performed.

- Acquiring Virtual Currency or Tokens – For this dimension, the KSC rewards users with virtual currency or tokens that are only valid within the community. The member can then exchange the token for goods available or in some cases for real currency. In the latter case, sites typically require a certain amount of virtual money to be collected before it can be exchanged.

- Ranking on Leaderboard – Some members like to be publicly recognized. Farzan et al. found that users wanted to measure themselves against other members and show that they were doing better (Farzan et al., 2008). In such scenarios, the use of leaderboards provides a way for users to compete to reach the top. Similar to visual games, leaderboards provide the listing of the top members who have accumulated a certain amount of points. Leaderboards either present ranking information as overall points or divided by categories to represent multiple rankings.

- Acquiring Activity Points – With this approach, members obtain points based on the execution of pre-defined activities within the KSC. Members can obtain points for accessing

the system, posting questions, or posting comments, among other activities. At the same time, members can lose points by asking questions ("Yahoo! Answers," n.d.). Farzan et al. modified a social network site used within the enterprise by adding activity points (Farzan et al., 2008). Their observations showed that after introducing the point systems, the level of activity increased. Members were interested on accumulating points as if it were a game. One important aspect to note about points is the notion of decay. Once members have achieved the highest possible points the effectiveness of the point system tends to decay. To keep users engaged, communities can add means for points to expire requiring users to maintain a certain level of engagement.

- Implicit work – This applies to members whose actual job is to assist other members via the KSC. For example, for communities that rely on library services, the librarian assigned to assist the member works at the library that also provides the service. Therefore, the motivation to participate on the community is implicit as part of the librarian's daily activities.

### 2.3.4   Communication Dimension

How KSCs communicate with their members is important because it maintains awareness of content. This communication can be asynchronous or synchronous. Asynchronous communication is the most common one in use, however, there are some situations where members might need to be quickly notified of posted notes (Richardson & White, 2011). For those scenarios, synchronous communication is used.

- Asynchronous – Most Online KSCs are asynchronous by design. In a sense, members post their questions on the knowledge community after which other members can read them when

they access their system or if they receive an email letting them know to check back on the website.   This approach might require askers to wait a usually undetermined amount of time to receive an answer.   Examples of these types of communities include Yahoo! Answers ("Yahoo! Answers," n.d.) and Stack Overflow (StackOverflow, n.d.).

- Synchronous – These KSCs use real-time communication channels such as instant messaging to allow questions and answers to reach intended answerers in real-time resulting in faster answers and faster updates on the answers (Richardson & White, 2011). When designing synchronous systems, there needs to be a consideration for the level of interruption that members face when receiving broadcasted questions (Hsieh & Counts, 2009).   Some examples of synchronous systems include Aardvark, Twitter Answers, Zephyr, IBM Community Tools, mimir (Hsieh & Counts, 2009) and IM-n-Expert (Richardson & White, 2011).

### 2.3.5   Question Types Dimension

It is important for KSCs to support different types of knowledge contributions on their sites. Most KSCs treat all contributions as a question when in fact there are many types of questions. Depending on the type of question, a different type of answer or response might be required. The following are the types of knowledge contributions that most knowledge communities support.

- Factual – The most expected type of knowledge request is that of a question asking for facts. In (Hsieh et al., 2010), the authors found that questions that ask for facts and advice tend to be the more sincere ones which correlate to a higher archival value but also urgency and

21

difficulty. Allowing members to specify whether they want a fact might result in shorter and better answers.

- Opinion – With this type of question, members are looking for an opinion rather than a fact about a topic. Opinion questions result in longer answers as well as a higher level of discussion and disagreement than a factual question. Different from factual questions, opinion questions cannot easily obtain a right answer but instead may yield several good points about a question as well as several personal opinion points (Agichtein, 2008).

- Advice – These questions are asking for advice. While different from opinion-based question in the sense that these questions can have an answer based on the experience of the responder, selecting the best answer for a question would require the answerer to address the points the asker included in the question. In fact, most askers might already know the answer to their question but want some level of assurance from a trusted source.

- Non-questions – These are questions that do not have a clear purpose or are just bogus in nature. Gingensohn and Lee (Girgensohn & Lee, 2002) noted that the type of incentive to participate might affect the number of non-questions received. For example, the number of non-questions may increase where members are required to contribute a minimum number of posts to be considered for a monthly raffle.

### 2.3.6 Moderation Dimension

One issue that KSCs have to address is how to control the posting of inappropriate content. For a community to be effective, it must have some order that members can follow. KSCs adopt different approaches when determining how to enforce order.

- Self-Regulated – Sometimes regulation can be counterproductive to the liveliness of a community. KSC members believe that for a community to be effective, it must not be too tightly regulated (Roberts & Roberts, 1998). Therefore, to allow freedom of expression, KSCs adopt a self-regulation approach. With this approach, the members themselves regulate their postings via unwritten rules.

- Moderated – To ensure that members post only appropriate and relevant material, a KSC might adopt a moderated model. This approach ensures that before any other member can see a post, a moderator needs to approve it. This approach works very well for specialty KSCs. However the moderator becomes the bottleneck for the availability of content. That is, postings would appear as soon as the moderator has found the time to review the posting and approve it. In addition, this approach might discourage some members to post as they might find it tedious to have to wait for their posting to appear. On the other hand, members can benefit from a level of moderation to ensure that postings that are not appropriate are flagged before they become public (Ardichvili et al., 2003).

### 2.3.7   Best Answer Selection Dimension

An important activity performed in a KSC is the identification of the best answer for a posted question. KSCs employ different approaches to select best answers. Most of the time, answer selection depends on the type of question posted. However, the criteria for what constitutes a "best answer" to a question are very difficult to determine. In Yahoo! Answers for example, a "'Best answer' feature allows the asker to choose  the correct answer (Fichman, 2011). However, what makes the asker knowledgeable enough to select the best answer? For this and other reasons, KSCs adopt different approaches to identify the "answer" to a question.

- Asker-selected - In this approach, the asker who posted the question is the only one who determines whether a best answer has been posted or not. We mentioned previously that Yahoo! Answers adopts this approach. However, market-based KSCs, where members pay to ask questions, also adopt this approach (Hsieh & Counts, 2009). After all, the asker is paying for an answer. The main limitation of this approach is that the selection of the best answer is left solely at the discretion of the asker.

- Community vote − In this approach, the KSC leave the identification of the best answer to community members who might vote on the answer they believe is the most appropriate. However, user ratings are problematic because they provide a subjective measure of answer quality. Users are not always knowledgeable enough about the topic of their own question and cannot accurately ascertain whether the answer is accurate or complete (Pera, 2011) (Fichman, 2011). This could lead to either an incorrect answer being accepted as the best answer or even if the best answer is provided, no one selects it for fear of being wrong. KSCs could turn to displaying the answers in sorted order to facilitate selection by users. For example, different metrics such as view counts, up/down votes, or other metrics could be used to bring awareness to potential best answers.

- Expert-provided − In this approach, an expert is the only person allowed to respond to questions. For example, Digital Reference Service and "Ask an Expert" systems assign questions to a specific person rather than open the floor to every member of the community to respond. While it is possible that a best answer is not available, if an answer is provided, it typically is a correct answer. Because of this restrictive process, some members have found ways to improve on the answers provided by experts by submitting comments on the answers ("Google Answers," n.d.).

- Hybrid - In a hybrid scenario, the KSC adds temporal constraints for selecting the answer. For example, some communities would give a predetermined time after an answer has been posted to allow the asker to determine whether the answer is accepted or not. If the time expires and no decision is made, the community could either unlock the question to be answered by other members (if expert driven) or accept the answer with the most votes as the correct one.

- Proportional voting – Sometimes selecting a single best answer is not feasible. An analysis of Yahoo! Answers found that answers to a question tend to complement each other by providing additional information with the most recent posts basing its facts on knowledge not already provided (Jain, Chen, & Parkes, 2009). In such a case, proportional voting allows multiple answers to be selected with all of them being proportionally correct.

- No best answer – Some KSCs do not need to select or accept a correct answer; therefore, they have no 'best answer' feature. Examples of such communities include WikiAnswers and the Wikipedia Reference Desk (Fichman, 2011).

### 2.3.8   Membership Scope Dimension

One of the most important assets of a KSC is its members and the membership directory. For this reason, a KSC needs to determine whether to share their members with other KSCs or to retain membership only within the realm of the community.

- Private - Once a user has built a reputation on a site, she will be reluctant to defect to a competitor since she would then have to build her reputation from scratch (Ardichvili et al., 2003). Communities benefit from a member lock-in to ensure they remain connected to their communities (Farzan et al., 2008).

25

- Shared – In a shared approach, the KSC shares the memberships with other KSCs or external services. The main advantage is that if a member has a membership with another external service, then the member would not need to remember another identifier to access that site. For example, in (Dearman & Truong, 2010) the authors examined Beehive, an enterprise social network site that allowed members to authenticate to the service using their corporate email addresses. Users were not required to create an account with the KSC thus ensuring that few barriers to entry are present. Other approaches to share memberships are through the use of OpenID or OAuth, which are becoming common on the web.

- Public – In a public scope approach, the community might not require any members to provide credentials to access the contents of the site. The contents of the site are public and therefore anyone who can reach the site can read its contents. On the other hand, in order to create content on the site, the KSC might require the user to authenticate to prevent bogus material, bogus evaluations, or other metrics such as votes from being added.

**Table 1.** Q&A systems and discussion forums grouped by dimensions.

| Dimension | Sample Values | Example |
|---|---|---|
| **Motivational** | Altruism | Ask.com[4], Knowledge Network (Ardichvili et al., 2003), Piazza("Piazza," n.d.), MicroBrowser |
| | Economic incentives | Yahoo_Answers, Portkey (Girgensohn & Lee, 2002) |
| | Gaming | Stack_Overflow, Portkey (Girgensohn & Lee, 2002), CHIPlace (Girgensohn & Lee, 2002), Answerbag ("Answerbag," n.d.) , Knowledge IN (Nam et al., 2009) |
| | Implicit Work | |
| **Member Reputation** | Badges | StackOverflow (StackOverflow, n.d.) |
| | Leaderboards ranking | Yahoo_Answers("Yahoo! Answers," n.d.), Answerbag |
| | Profile | CHIPlace (Girgensohn & Lee, 2002) |
| | Total points | Yahoo_Answers, StackOverflow, Answerbag |
| | Authority or credibility | Knowledge Network (Ardichvili et al., 2003) |
| | Number of best answers | |
| **Question types** | Factual | MicroBrowser |
| | Advice | Piazza ("Piazza," n.d.), MicroBrowser |
| | Opinion | |
| | Non-question | |
| **Membership Scope** | Shared | Yahoo_Answers |
| | Private | Portkey(Girgensohn & Lee, 2002), Knowledge Network (Ardichvili et al., 2003), MicroBrowserr |
| | Public | CHIPlace (Girgensohn & Lee, 2002), StackOverflow |
| **Answerer Selection** | Expert | Google Answers("Google Answers," n.d.) |
| | Digital Reference Service | |
| | Community | Stack_Overflow (StackOverflow, n.d.) |
| | Member Matching | IM-an- Expert (Richardson & White, 2011) |
| | Question Matching | QAR (Pera, 2011) |

---

[4] http://www.ask.com

| Communication | Synchronous | Mimir (Hsieh & Counts, 2009), IM-an- Expert (Richardson & White, 2011) |
|---|---|---|
| | Asynchronous | CHIPlace (Girgensohn & Lee, 2002), MicroBrowser |
| **Moderation** | Moderated | |
| | Self-regulated | Yahoo_Answers, Stack_Overflow |
| **Best Answer Selection** | Expert provided | Google Answers |
| | Community vote | Stack_Overflow |
| | Asker Selected | MicroBrowser |
| | No selection | WikiAnswers, Wikipedia Reference Desk (Fichman, 2011). |
| | Hybrid | |
| | Proportional | |

## 2.4 FORUM VISUALIZATION

At any point in time, thousands of discussions are occurring online on a vast range of topics, from how to change oil on your car to political science discussions. Most of these discussions occur on discussion forums and Q&A systems that are for the most part text-based, making them easy to access and easy to use. The use of text-based browsing schemes make it difficult to obtain insights among all discussions at the same time and they tend to give the same space to each discussion thread with no regard to attributes that might be interesting to forum users. Additionally, discussion entries are typically arranged in chronological order with most recent discussions appearing first. In this interaction model, important or interesting discussions are not necessarily the most recent ones; there might be helpful discussions that have been created in the past (Indratmo & Gutwin, 2008). Exploratory tasks and finding insights in discussions becomes tedious when users browse a large collection of discussions, making them unwieldy to navigate and review.

Visualization provides a powerful means of making sense of large collections of data. The power of visualization techniques is a consequence in large part from their ability to reduce mental workload (Sebrechts & Cugini, 1999). By mapping data attributes to visual properties such as position, size, shape, and color, visualization designers leverage perceptual skills to help users discern and interpret patterns within data. For visualizations to be effective, they need to allow users to perform meaningful visual analysis on the data. Using single static images might provide answers to, at best, a handful of questions. Instead, visual analysis allows users to

iteratively perform a process of view creation, exploration, and refinement. Performing this analysis process in an iterative manner allow users to explore the data and develop insights about significant relationships, domain-specific contextual influences, and causal patterns (Heer, Shneiderman, & Park, 2012).

Studies on social interaction aim to understand the social dynamics that occur among members of a discussion forum. Although a number of visualization systems have been designed to process discussion forums, not all follow the same approach or share the same goal. Forum visualization systems have been designed with a goal to perform analysis of forum activity (Wanner et al., 2011)(Trampuš & Grobelnik, 2010), to study the social interaction and semantic visualization of forum participants (Gibbs, Olexa, & Bernas, 2006) (Xiong & Donath, 1999)(Perry & Donath, 2004)(Bratitsis & Dimitracopoulou, 2006)(Gibbs et al., 2006), to support navigation of large scale forum data (Giguet & Lucas, 2009)(Wattenberg & Millen, 2003)(Engdahl, Köksal, & Marsden, 2005), to support the use of discussion forums in education domains (Mazza & Milani, 2005)(Mazza & Dimitrova, 2005)(Jyothi, McAvinia, & Keating, 2012) and to study participants using emotions (Donath, Karahalios, & Viegas, 1999) and even forum activity using motion (Lam & Donath, 2005).

Forum analysis visualization systems integrate the use of content analysis with visual analysis to explore and search discussion forum content using various features. Wanner et al. (Wanner et al., 2011) designed an exploratory forum analysis system, ForAvis, to help users analyze user-generated content in online forums. ForAvis presented an overview of the discussions threads and was also able to show structural levels of user forums. The system made use of a flexible visualization that allowed users to perform forum analysis by changing the ordering and mapping of colors of discussion threads. Fortuna et al. (Fortuna et al., 2005)

provided a general approach to visualize text document collections using Latent Semantic Indexing to extract background knowledge from a corpus of text documents. The authors then showed how their technique can help in data analysis, for instance, for finding main topics that appear in larger sets of documents using a visualization system. Trampus and Grobelnik (Trampuš & Grobelnik, 2010) analyzed discussion topics and trends in discussion forums to deduct public opinion from political and social science discussion forums. The system consisted of three visualizations: a visual browsing suggestions mechanism, a semantic 'atlas' providing thematic overviews of large forum content, and a timeline displaying temporal evolution of forums. Lee (Y. Lee, 2012) proposed a smart visualization system, IssueGalaxy, that extracts current issues or topics from online forums. The system uses a metaphor of a galaxy with the current issues represented as suns and planets around the suns representing related articles. Using this metaphor, users could analyze how similar two or more issues that are current were by the spatial placement of the suns and their planets in the galaxy.

It is difficult for participants in online discussion forums to form a clear mental image of other participants. Visualizing how participants behave, participate and interact in the discussion forum can help not only other participants but also forum moderators to build this mental image of other participants and suggest how they should engage with them. Social indicators also act as an additional motive for a user's activity, fostering social interaction leading to a collaborative online environment. Gibbs et al. (Gibbs et al., 2006) designed Mapping Temporal Relations of Discussion Software (MTRDS) to look at communication patterns and online behaviors in asynchronous communication using their temporal data to represent them as maps to illustrate behavior of online discussants. MTRDS facilitated group discussions, for example by allowing users to see if discussions were interactive or one-way monologues. Xiong and Donath (Xiong

31

& Donath, 1999) instead used a garden metaphor to visually represent message board participation. A forum was represented as a garden with participants represented by a flower. The height of the flower encoded the involvement length of the participants in the forum while petals represented the posts made. The more participants have posted in the forums, the more petals there are. A forum with little activity would appear as a neglected garden, with mostly stems and little flowers. In another study, Perry and Donath (Perry & Donath, 2004) used anthropomorphic visualizations to present historical information about participants in online spaces. Anthropomorphic visualizations use the human form as the basis for the visualization with various data about an individual's online behavior mapped to different parts of a body.

Visualization tools appear to offer potential in providing a practical overview of discussion threads easily. One of the important challenges of the concept of visualization is to present as much information as possible in a given display area without overloading either the display, or the cognitive capability of the person viewing. Giguet and Lucas (Giguet & Lucas, 2009) designed Anagora, a graphic tool for identifying collaborative work by tracing discussion threads along a time axis. Anagora selected the best resolution for a forum to fit on a screen by choosing the most appropriate time scale (days, decades, months or more) according to the forum contents. In another study, Engdahl et al. (Engdahl et al., 2005) worked on visualizing threaded discussion forums on compact displays using squarified treemaps. Using treemaps, discussion threads in discussion forums were rendered as colored rectangles, thereby using 100% of the limited screen space. The authors reported a significant improvement in time for finding the largest and most active threads. The benefits derived from using treemaps on desktop computers were valid for small screens such as mobile devices. Wattenberg and Millen (Wattenberg & Millen, 2003) focused instead on visualizing individual large-scale discussion threads using a

system called Conversation Thumbnails.  Conversation Thumbnails used a focus+context visualization of message-level metadata to provide an easily navigable overview of a discussion.

With the increased use of computer-mediated communication tools in educational settings and in particular, the use of discussion forums for interaction between students and instructors, there is greater need to understand how students are making use of discussion forums.  During face-to-face class discussions, instructors are able to observe the verbal and non-verbal behaviors of students, which can help them efficiently and effectively manage a conversation.   Online conversations lack the visual and verbal cues of face-to-face communication such as gestures, facial expressions and direct dialogue resulting in more complex management of conversations.  This leads to a common problem in distance learning: the monitoring and checking of students' activities in courses delivered with distance education tools.  Mazza and Milani (Mazza & Milani, 2005)  used graphical representations of student tracking data collected by learning tools to help instructors become aware of what is happening in distance learning classes.  Instructors gained insights about student participations through an overview visualization of all the discussions in which students participated. For each student, the visualization indicated the number of messages posted (using a square), number of messages read (using a circle) and finally the number of threads started by the student in the discussions (using a triangle).   Mazza and Dimitrova (Mazza & Dimitrova, 2005)  presented CourseVis, a system that generates graphical representations of multidimensional student tracking data. CourseVis assisted instructors with getting a better understanding of social, behavioral, and cognitive aspects of students in their class.  Using CourseVis, instructors can identify tendencies in their classes and quickly discover individuals that need special attention, allowing the instructor to provide better support to their students. CourseViz visualized discussion threads

along a timeline indexed by their creation date. Users could order discussions by either author or topic. Size and color coding was used to represent the number of replies or follow-ups for each discussion thread. Jyothi et al. (Jyothi et al., 2012) described a visualization tool called VIMS to aid the analysis of online communication on a day-to-day basis by teachers or forum moderators to review the development of a discussion and to support appropriate interventions. VIMS supported the analysis of interactions following Schrire's classification schemes. Messages were visually depicted as circles, with instructor messages differentiated from students' messages by different color schemes. Connections between circles represented flow of information with an arrow representing the direction of the interaction.

Forum visualization has also been used to capture emotional content. Donath et al. (Donath et al., 1999) looked at social and scalability aspects of forum visualizations. The Loom project aimed to provide a visual interface for browsing newsgroups archives to discover social patterns often obscured in text-only interfaces. The Loom project developed visualizations that provided a visual thumbprint of each group to quickly establish the atmosphere of each group by identifying the mood of participants. Mood detection was performed by doing text analysis and classifying discussions into one of the mood categories using a simple decision-making algorithm. Classified discussions where then color-coded and displayed on an overview graph of a forum.

Forum visualization was not limited to static or manipulable visualizations, but work has also been done to use motion. Motion provides a useful dimension to visualize large-scale data and activity because it is perceptually rich and catches our attention rapidly. Lam and Donath (Lam & Donath, 2005) designed Seascape and Volcano to visualize online discussions using motion as the strongest visual appeal to attention. The authors concentrated on visualizing the

most prominent characteristics of a newsgroup such as size, activeness, and recentness. Using motion, inactive threads form slower waveforms while active threads propagate across the screen faster.

## 2.5    DESIGN PATTERNS

As mentioned in Section 1.1, the second objective of this work is to determine whether the introduction of design patterns in Q&A systems increases student participation and peer-to-peer learning by facilitating knowledge discovery, generation and reuse.

### 2.5.1   Patterns History

In the 1970s, Christopher Alexander (Alexander, Ishikawa, & Silverstein, 1977)(Alexander, 1979) wrote a number of books documenting the notion of patterns in the Architecture field to facilitate design and construction of towns and buildings. In his introductory talk about patterns, Alexander (Alexander, 1999) described that in architecture the mass production of living structures was not the only objective. Instead, he described that it was also important to be able to adapt the living structures to the local environment. As mass production of housing developments began to take hold, it was important to be able to bring this adaptation into a particular situation to facilitate a good environment. Alexander and his colleagues developed 253 patterns for building and planning towns, neighborhoods, houses, and other livable spaces (Alexander et al., 1977).

The use of patterns was popularized in software engineering by Gamma et al. in their book *Design Patterns: Elements of Reusable Object-Oriented* Software (Gamma, Helm, Johnson, & Vlissides, 1994). Borrowing on the idea of design patterns from Christopher Alexander, the authors set out to document recurring designs from numerous software projects and documented them in their book for reuse.

## 2.5.2   Defining Patterns

The Merriam-Webster dictionary defines patterns as "something designed or used as a model for making things" ("Merriam-Webster Dictionary," n.d.).   In Alexander's view, "… each pattern describes a problem which occurs over and over again in our environment, and then describes the core of the solution to that problem, in such a way that you can use this solution a million times over, without ever doing it the same way twice" (Alexander et al., 1977).  Thus, patterns enable us to document a known recurring problem and its solution in a particular context, and to communicate this knowledge to others.   A key element in this description is the word recurring, since the goal of the pattern is to foster conceptual reuse over time.  A quality of patterns is that they are neither too general nor too specific.

In its basic form, patterns are composed of a Name, Context, Problem, Solution and related patterns.  Name identifies the design pattern and distinguishes it from other design patterns.  Context refers to the particular situation or state in which the activity takes place.  It is the environment, surroundings, situation or conditions within which something occurs. Problem refers to the challenge, which occurs when the activity is in the particular context. It is an unsettled question that requires further investigation and an answer.   Solution is a repeatable

approach to solve the problem when it occurs in the given context. Finally, related patterns are those that have some association or dependency to the design pattern.

Does having a solution to a problem in a context qualify as a pattern? Not necessarily. The important fact to understand is that the solution to the problem is generic in the sense that it can be applied to multiple situations with slight modifications based on the context.

### 2.5.3   Pattern Language

When describing patterns, Christopher Alexander placed emphasis on an entire language for design, since the usefulness of patterns was not only in providing solutions to common problems, but also in seeing how they related and affected one another (Alexander et al., 1977). Design patterns are a shared language. A collection of one or more patterns in a particular domain form a pattern language as shown in Figure 1. Patterns that form a pattern language should include their relationship to other patterns and to the entire pattern language.
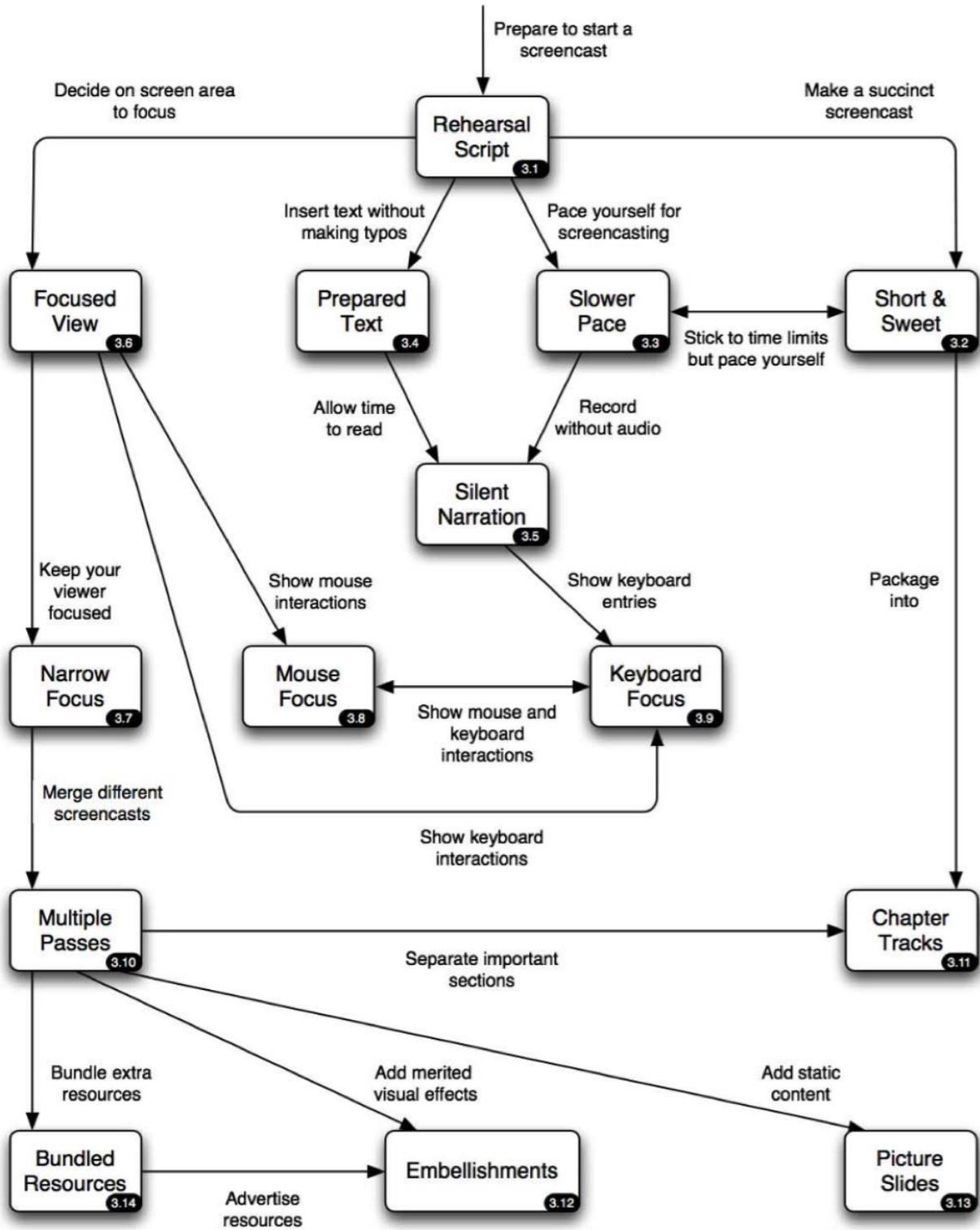
**Figure 1.** A pattern language for screen casting (Cook, Boness, Bartsch, Harrison, & Wernick, 2005). Connections represent the recommended flow when making a screencast and the applicable patterns for each step.

### 2.5.4   Adopting Design Patterns

Patterns have a natural life cycle.  They emerge from lessons learned in a particular domain which are then adapted by domain experts (Petter, Khazanchi, & Murphy, 2010).    The use of patterns has had a wide adoption in different domains with several pattern languages being defined.    Given the steadily growing number of pattern languages and approaches for using design patterns in the literature, it is hard to be aware of all the patterns that exist and to select appropriate patterns for problems at hand (Durdik & Reussner, 2013).

The work from Gamma et al. (Gamma et al., 1994) resulted in the extensive adoption of patterns in object oriented programming and software design in general.  The main benefits were the reuse of best practices and improvement in the quality of software systems.  By using patterns, software engineers learn from the experiences of experts and design new ideas and materialize concepts by following guidelines defined in the patterns.

Several approaches have been made to improve the adoption and facilitate the use of design patterns.  With Model-based design and architecture (Szekely, 1996), developers define systems at a higher level of abstraction improving the quality and reducing the cost of writing code.  Unified Modeling Language (UML) is a common modeling language and it facilitates the modeling of applications and the implicit integration of design patterns(Dong, Yang, Zhang, & Member, 2007).  UML models are able to capture organization and behavior of code in an abstract way and have become a common language for the documentation of software design patterns.  This level of abstraction facilitates code documentation making it easier for new programmers to learn the code behavior.

However, while model-based design and architecture helps software architects to design the system, it makes it harder to translate the models into actual working code, particularly for User Interface (UI) programmers (Lin & Landay, 2008). UI programmers lack the programming skills required to implement complex systems and at the same time, using software design patterns do not provide good support to capture human-computer interaction and related user interface elements at a finer detail. To address this, the use of patterns in UI design resulted in the documentation of proven UI best practices and reusable ideas which aim to address particular situations of human-computer interaction and behavior (Tidwell, 2005)(Lin & Landay, 2008)(Deng, Kemp, & Todd, 2005). For example, Figure 2 shows a sample UI pattern for organizing and displaying elements.



**Figure 2.** Two-panel-selector UI pattern (Tidwell, 2005)

The success of patterns in UI design and its notion of abstraction have led to new opportunities. With increased interest in the design of cross-device UIs, designers are looking to the use of patterns to address problems in designing interfaces that work in multiple devices Lin

40

et al. (Lin & Landay, 2008) designed Damask, a tool that uses patterns to build web UIs that run on PCs and mobile phones as well as voice-guided interfaces. As can be seen, the use of patterns to build UIs is not new but it is in fact growing as new user interfaces and interaction modalities between human and computers evolve.

Design patterns have also been incorporated into educational settings. Astrachan et al. developed instruction material and frameworks that incorporate design patterns to facilitate teaching object oriented programming to Computer Science students (Astrachan et al., 1998). Muller (Muller, 2005) proposed Pattern Oriented Instruction as a pedagogical technique that recommends adopting design pattern in course instruction with a goal of improving problem-solving skills in students. Muller et al. (Muller, Ginat, & Haberman, 2007) also showed that teaching algorithmic patterns as part of an introductory Computer Science course helped students to construct solutions to problems by being able to decompose the problem into sub tasks and to associate algorithmic patterns to solve the sub tasks. More recent work has focused on adopting design patterns in teaching parallel and distributed programming to undergraduate students (Brown et al., 2014). Wilkinson et al. described an approach of using computational patterns to help students solve parallel programming tasks by identifying relevant computational patterns first and before writing the implementation (Wilkinson, Villalobos, & Ferner, 2013).

Patterns have also been defined to assist instructors in preparing instructional material. Pedagogical patterns consist of successful techniques for teaching and learning collected by experienced teachers to facilitate transfer of their learned knowledge to new instructors (Bergin, 2012). Carle et al. presented how novice instructors benefited from the annotation of learning material with references to pedagogical patterns to design course content by leveraging the best practices (Carle et al., 2007).

41

In this dissertation, our focus is to introduce the concept of design patterns to facilitate the discovery and generation of new knowledge by students. Using design patterns, students can reuse the approach followed to answer similar questions in order to generate a new solution. When viewing questions in a more abstract way, students can find more similitudes among them resulting in re-occurrence. Following the solution suggested by an appropriate pattern, a student can provide an answer, in a more concrete form, to the current question.

# 3.0    MOTIVATING PARTICIPATION IN EDUCATIONAL SETTINGS VIA AN

# IMPROVED Q&A SYSTEM

## 3.1    INTRODUCTION

Graduating Undergraduate students with a Science, Technology, Engineering and Mathematics (STEM) degree face challenges for which they might not be prepared once they enter the workforce. The challenges include open-ended and ill-defined problems not typically addressed during a normal course of education. A skill that is helpful with addressing these problems is research exploration commonly performed with close faculty member involvement. However, the mentoring time required from the instructors compounded with the number of students makes this approach not scalable. In a study conducted by Smith et al. (Smith et al., 2009), the authors found that peer discussion enhances learning even if none of the students knew the answer originally. One technique that fosters peer learning and knowledge sharing among students is to use online knowledge sharing systems such as Question and Answer (Q&A) systems. This approach serves three main purposes; first, it eases the fear of tackling unknown problems which can lead to discouragement. Second, it provides students with collaboration and knowledge generation skills that are highly beneficial in their post educational careers. Third, it reduces the involvement time required from the instructor.

The increased use of the Web as a knowledge repository and social medium has led to the proliferation of online knowledge sharing systems. These systems rely on their members to create the knowledge by either following a discussion thread or question and answer format. The knowledge created becomes available to all members of the community. Following a Question and Answer format becomes more intuitive to users. When users have a need to know something about a topic, it is easier to represent their inquiries in the form of a question and ask them to a trusted source rather than using a search engine and filtering on the results.

In this section we describe MicroProbe (Figure 3), an improved Q&A system for educational settings that integrates gamification concepts to motivate student participation. We performed a pilot study in two courses at a local university. Among the major observations, we noted that graduate students tended to participate more with knowledge generation while undergraduate students performed more knowledge discovery tasks. Survey data also showed that while students embraced the idea of acquiring points for contributing knowledge, there was reluctance from students who felt as if they were competing against each other, especially when a leaderboard was used.
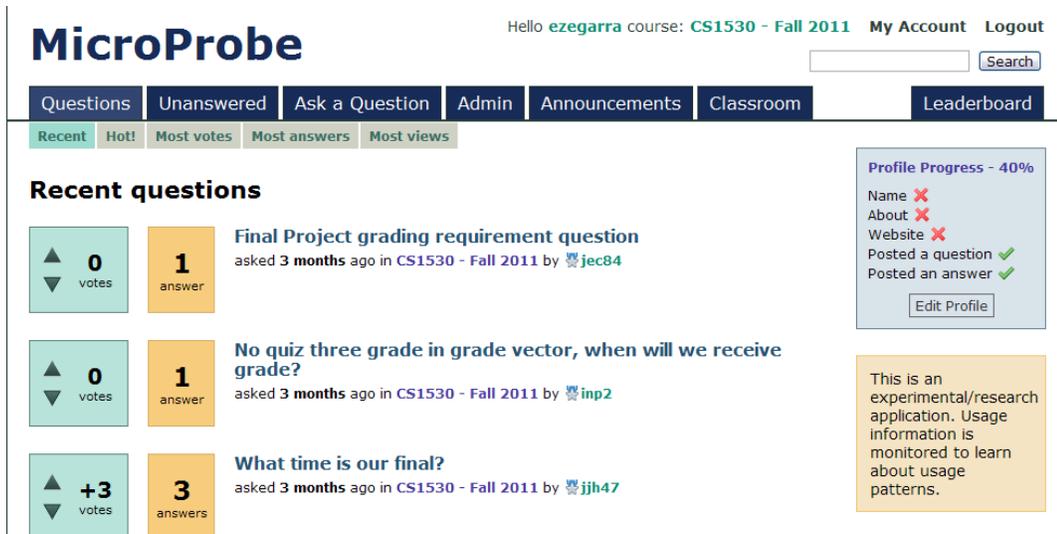
**Figure 3.** Main User Interface of MicroProbe. The tabs at the top allow users to navigate to different functionality. Gamification concepts such as points and leaderboards were added to motivate participation.

## 3.2    USING GAMIFICATION TECHNIQUES TO MOTIVATE PARTICIPATION

For any knowledge sharing system to be effective and thrive, even in an educational domain, it needs to ensure that its members are constant participants. Without participation, either active or passive, their perceived value decreases (Ardichvili et al., 2003). Therefore, it is important to be able to address the motivational aspect to bootstrap participation.

The common expectation is that as the number of members engaged with the system grows, the participation level increases resulting in larger number of questions and answers posted. Raymer (2011) defines engagement as "occupying the attention or efforts of a person". He goes further to describe that for something to be engaging, members must like it and want it. For example, video games have an intrinsic capability to motivate users to remain engaged in an

activity with unparalleled intensity and duration.   Most people can associate themselves with either liking or wanting to play games.  Gamification is a method to improve user experience and engagement by adopting techniques used in video game design (Raymer, 2011).  While the term gamification might imply that users are playing games, in fact it is the adoption of feedback mechanisms typical in game designs that enables the exploration of features of the system (Deterding, Dixon, & Khaled, 2011).   Therefore, by adopting the techniques used in game design, Q&A system designers can make the experience more enjoyable and indirectly engage its members to participate.

Q&A systems have adopted several gamification techniques in order to motivate usage. StackOverflow members obtain points for answering questions and receive points for good answers.  In (Farzan et al., 2008), the authors found that giving more points for certain features motivates participants to use them.  Gamification techniques can also stimulate participation by allowing users to compare against each other.   In Answerbag, Yahoo Answers and StackOverlow, users accumulate points and are shown in a leaderboard where they can compare themselves against other users.

In MicroProbe, we similarly include the use of gamification concepts such as points and leaderboards.

### 3.3    MICROPROBE SYSTEM OVERVIEW

Our Q&A system, called MicroProbe, was based on a modified version of the Question2Answers platform.   The base platform was extended to support the notion of courses allowing users to

switch between course pages (if applicable) and only see questions posted for their particular course. Upon login, students were presented with the landing page (Figure 3). The landing page also provided access to the main features of the system through a tabbed layout. From this page, students can navigate to the different sections of the system. The *Questions* tab presented a list of questions posted to the course. The *Unanswered* tab listed questions that had not received a response. The *Ask a Question* tab allows students to post a question to the course. The *Admin* tab, visible only to administrators, controls the behavior of the system. The *Announcements* tab lists course announcements of general interest to students, posted by the instructors. Finally, the *Leaderboard* tab shows a ranking of students based on activity, knowledge contribution and participation.
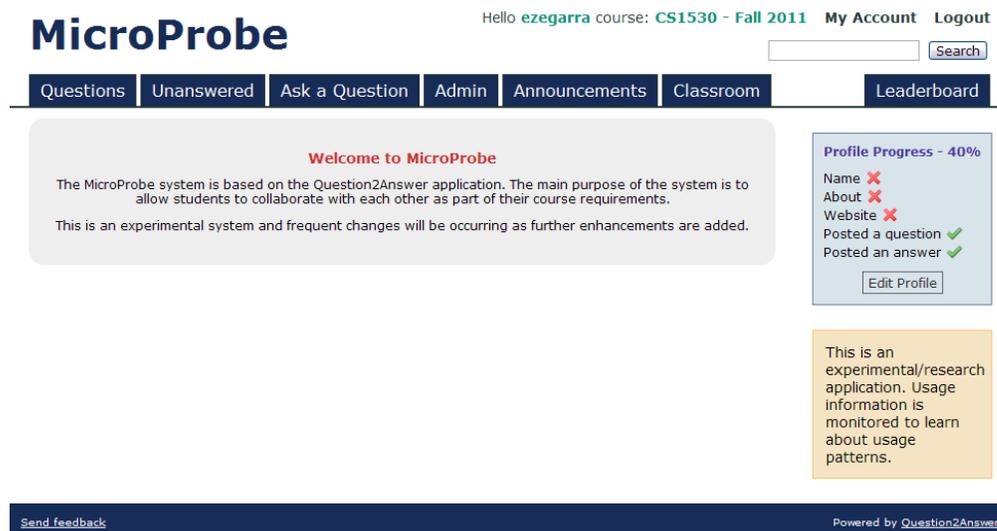


**Figure 4.** Landing page. The first time users enter the system they are presented the landing page view. The Profile progress view on the right identifies items that students have not completed in their profile

### 3.3.1    Asking Questions

When students wanted to ask a question in the course, they selected the Ask a Question tab and were presented with the form shown in Figure 5. The form requested users to provide a question title and a rich text formatted question text. Students had the option of being notified whenever another student either commented or submitted an answer to the questions.  This feature was helpful in prompting users to come back to check the system.  Once the question was submitted, students could navigate to the question and view its details including any answers provided to it.

**Ask a question**

**Your question in one sentence:**

**Category:**
CS1530 - Fall 2011

**More information for your question:**

☑ Email me (ezegarra@cs.pitt.edu) if my question is answered or commented on

Ask the Question

**Install required?**

+4 votes

In the acceptance testing checklist, it says that the we must first "Install Social Network Service Center (SNSC) software." for the demo.

Because our project is a website, the user cannot "install" anything, so I assume we can skip this step?  The project is running on the CS web servers.

asked **2 months** ago in **CS1530 - Fall 2011** by **djc42**

edit    flag    hide    answer    comment

**1 Answer**

0 votes

you are right.  you can skip this step.

answered **2 months** ago by **changsk** Gold

edit    flag    hide    ask related question    comment

**Figure 5.** Ask a Question form and question details.  Students provided a question title and the text of the question. Once a question was posted, students could view the question and any answers for it.

### 3.3.2 Viewing available questions

If a student wanted to see what questions were available, he could select either the *Questions* or the *Unanswered* tabs. Upon selecting the Questions tab the student was presented with a list of the most recent questions posted for the course (Figure 6). The results displayed the title of the question asked, the number of votes it received, the number of answers it received, how long ago it was asked and by whom. To view the details of a question, a student simply clicked on its title. Selecting the Unanswered tab would in turn only show those questions that have received answers but do not have an accepted answer or have not received any answers.



**Figure 6.** Question listing view. The view presented the most recent discussions either created or updated. Each result contained the number of votes, answers, the title, the author and the course it was asked on.

### 3.3.3 Course Announcements

As part of the design process, we collected feedback from the course instructor to ensure proper functionality was in place. One such requirement was the desire to post announcements to the class without following the defined Q&A format. Some examples of announcements would be if the instructor wanted to communicate important dates such as midterms or quizzes or clarifications about a particular homework. To address this, we decided to add the *Announcements* functionality. Only instructors and Teaching Assistants were able to create announcements; however, all users of the system could see the announcements. Users were able to access the announcements by selecting the Announcements tab from the main navigation tabs.

### 3.3.4 Encouraging student participation

One difficulty with introducing a new tool to a classroom setting is getting students motivated to participate and use the system. To address the motivation aspect, three approaches were used. First, the instructor suggested that students post questions of interest to the entire class using the system. This way, rather than providing the answer to each student individually, the instructor had the option to reply to all students at the same time. The second approach consisted of the use and integration of gamification concepts into the design of the system. The final approach was the use of notifications of updates such as new questions or announcements.

When accessing question details or viewing a listing of questions, students could view the points other students had acquired. MicroProbe awarded students with points based on activities performed. There were four dimensions monitored: Usage Activity, Knowledge Contribution,

Participation and Total Points. Students were awarded points under the Usage Activity just for using the system, for example, logging in, viewing the questions or answers, updating their profile, etc. The Knowledge Contribution dimension awarded points whenever a student answered a question using the system. The Participation dimension awarded points for asking questions or voting on the questions or answers provided. Finally, the Total Points dimension summed up the points across all three dimensions.

A common gamification concept is the use of leaderboards. We made use of the leaderboards to present the points students had acquired for the different dimensions. The motivations for students to participate rely on the fact that students would want to be ranked on top of the leaderboard.

To keep students informed about activity in the system, we added support to notify students, via email, whenever a new question was posted, if the instructor posted a new announcement and also whenever an answer was submitted to one of their questions. In (Girgensohn & Lee, 2002), the authors noted spikes in participant activity in discussion forums whenever email notifications were sent; thus we expected to achieve the same behavior with our email notifications.

One additional feature available in MicroProbe was the profile completion progress. The goal of this view was to motivate students to complete their profiles. The view showed a profile completion percentage and listed the missing and completed information.
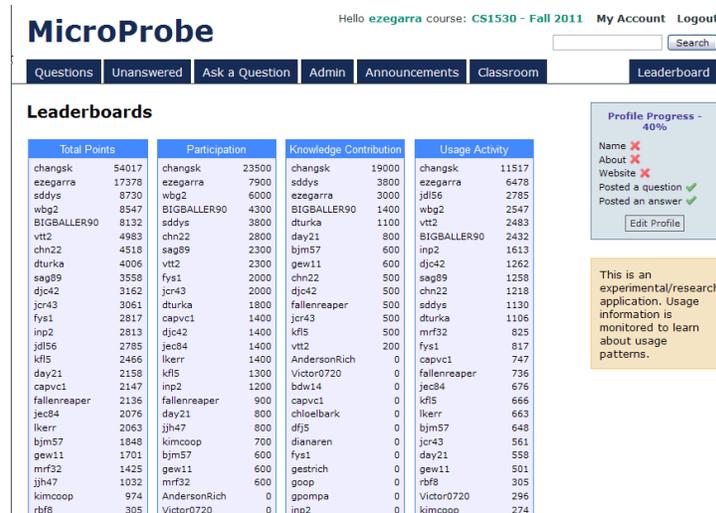
**Figure 7.** Leaderboard. Four different dimensions were used to rank students: Total points, participation, knowledge contribution and usage activity.

## 3.4    PILOT STUDY EVALUATION

In order to understand the motivational aspects of students' participation and the use of Q&A systems in educational settings, we conducted a pilot study. The objectives of the study were to find out:

- What motivates students to ask questions?

- What prevents students from asking questions?

- What motivates students to answer questions?

- What prevents students from answering questions?

- Is the system easy to use?

### 3.4.1 Methodology

The system was introduced to students in two courses: CS1530 – Software Engineering and CS2310 – Multimedia Software Engineering. CS1530 was an undergraduate course while CS2310 was a graduate level course at a local university. In order to begin the bootstrapping process to get the students to use the system, the instructor requested students to post a sample question and answer as part of their first assignment. This step was important as it helped the students become familiar with the system early. Additionally, the instructor suggested students to use the system and post any questions about the class directed to the instructor or the Teaching Assistant so that other students were aware of answers supplied. After this, students' use of the system was at their own discretion. Both the instructor and the Teaching Assistant monitored the system for questions from students.

At the end of the semester, a survey was administered to students. The survey consisted of open-ended, multiple choice, and Likert-scaled questions. In addition, several demographic variables were recorded such as age, gender, education years and degree program. In addition to the survey, the experiment collected and stored server data such as login sessions, page visited, actions performed, and contents of the questions and answers for later analysis. Only those students that were present at the beginning of the last session took part in the survey.

### 3.4.2 Results and Discussions

### 3.4.2.1 Participants

A total of 43 responses (71% male, 29% female) were collected using the questionnaire among students from both the graduate and undergraduate courses. The respondents included 31 undergraduate and 12 graduate students. Twenty-two out of 31 of the undergraduate students and 11 out of 12 graduate students reported Computer Science as their major. Six undergraduate students reported Computer Engineering as their major. The age distribution by course is presented in Figure 8.

We used Google Analytics to monitor page tracking data for our system. From our data collected, we noted that the majority of students accessed the system using a desktop computer, with Google Chrome being the preferred browser. We also noted students accessed the system using a browser on their mobile device. Among the students that used their mobile devices, Android and iOS were the most popular operating systems.

### 3.4.2.2 Motivators and inhibitors to asking questions

Undergraduate students reported a higher level of participation than graduate students by asking questions in the system. Fifty-eight percent of the undergraduate students participated by asking questions while only 33% of the graduate students did. As an explanation of why graduate students asked less questions, a student indicated that usually graduate students tend to work on different projects on their own so they preferred to solve problems individually rather than asking other students who might not be working on similar problems for help.
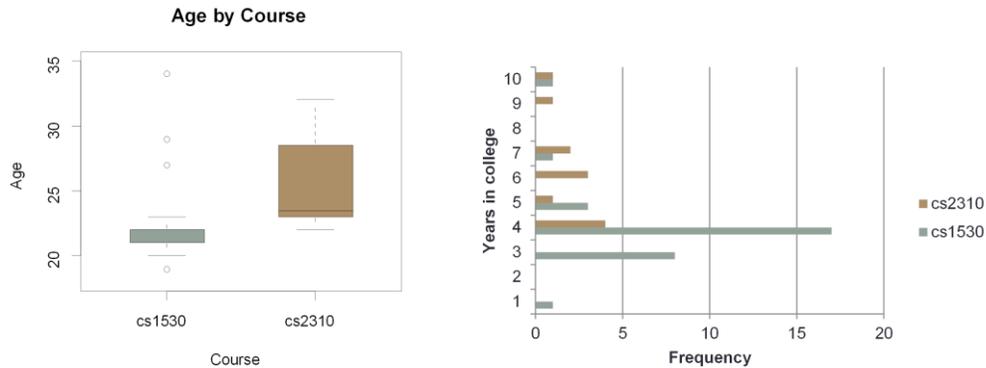
**Figure 8.** Participant demographic information. CS1530 – undergraduate, CS2310 – graduate. The graphs show first the age distribution of students by course and years of college experience.

The main reason students indicated for asking questions was to get feedback about class materials (undergraduate 90% and graduate 66%). In particular, 51% of undergraduate and 33% of graduate students wanted to get clarifications on homework assignments, and 25% of both graduate and undergraduate students wanted help figuring out technical details about an assignment. Regarding the use of points, only 3% of undergraduate students and none of the graduate students reported being motivated by getting points to increase their ranking on the leaderboard.

When looking at the inhibitors to asking questions, we grouped students' responses into three groups: privacy, social loafing and gaming. For the privacy aspect, students preferred to interact with their own group or project members (undergraduate 25%, graduate 8%) and were also afraid of asking the question to the entire class (9% undergraduates, 25% graduates). Social loafing is the condition where students preferred for others to do the work (Karau & Williams, 1993). Students reported this behavior by waiting for other students to ask questions they were

interested in (19% undergraduates, 16% graduates). For the gaming perspective, students were expecting some reward from their participation such as grade or points for asking questions (6% undergraduates, 8% graduates). This result was interesting to note because while graduate students claimed not to be motivated to ask questions by gaming concepts, 8% claimed that the absence of a reward prevented them from asking questions.

### 3.4.2.3 Motivators and inhibitors to answering questions

Graduate students reported a higher level of participation for generating new knowledge in the form of answers (23% undergraduate, 42% graduate). When looking at what motivates students to respond to questions, we found two major trends: altruism and gaming. Altruism is the selfless reason to help others without expecting any reward. Among student responses, graduate students reported the higher level of altruism (51% undergraduates, 83% graduates). On the other hand, undergraduate students reported being more motivated by gaming techniques such as participation points or raising their rank in the leader board (22% undergraduates, 0% graduates) while none of the graduate students did.

The most cited reasons that prevented students from answering questions was not knowing the answer (35% undergraduates, 66% graduates) and being afraid of answering the question incorrectly (35% undergraduates, 25% graduates). Similar to the motivators for answering questions, 9% of undergraduate students reported that they did not answer questions because they were expecting some reward (9% undergraduate, 0% graduates). Lastly, 9% of the undergraduate students reported that they expected the instructor or the TA to answer the questions. This behavior was not reported by graduate students, and we believe it was because graduate students were less reliant on the instructor.

**3.4.2.4 Using MicroProbe in a classroom setting**

As reported by (Huang et al., 2014)(Mamykina, Manoim, Mittal, Hripcsak, & Hartmann, 2011),

usage of the system reported by students could be characterized by the presence of a few power

users frequently visiting the system and browsing questions and answers, while some students

rarely accessed the system, sometimes even less than once a month.  Survey results found that

the majority of students agreed that using the system was an effective alternative to going to the

TA or instructor (Figure 9).  Also, most students agreed that being able to read the answers from

other students was helpful and informative (Figure 10).  These results suggest that using a Q&A

system in an educational setting facilitates peer-learning and knowledge sharing.

| | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|---|
| cs1530 | 6.90% | 10.34% | 31.03% | 41.38% | 10.34% |
| cs2310 | 8.33% | 0.00% | 33.33% | 25.00% | 33.33% |
| combined | 7.32% | 7.32% | 31.71% | 34.15% | 17.07% |
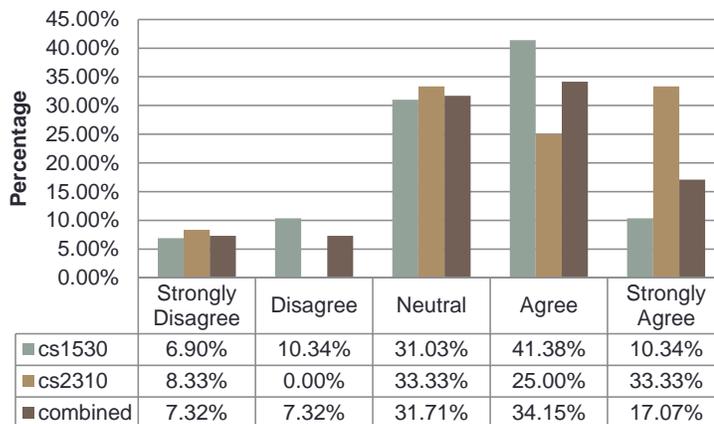
**Figure 9.** Is asking questions using the system better than going to the TA or to the professor?
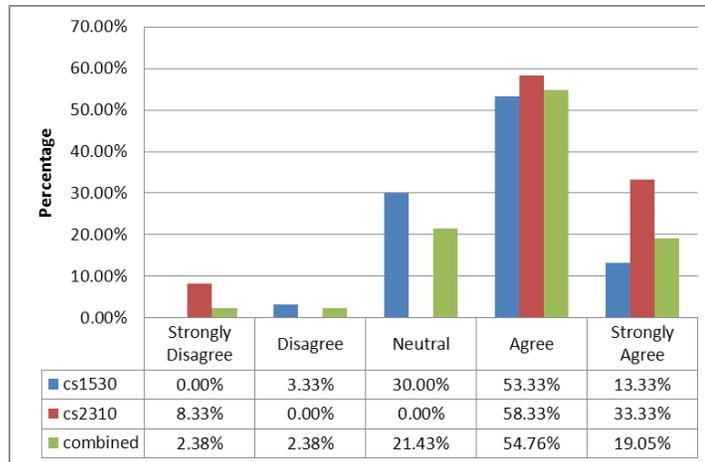
**Figure 10.** I feel it was helpful/informative to read answers given by other people

### 3.4.2.5 Use of gamification concepts

We got mixed results on the use of gamification concepts. Similar to the responses received for motivators to asking and answering questions, students were not as highly motivated by the use of points for posting questions (Figure 11) or to answer questions (Figure 12) as we had expected. However, for answering questions, survey responses show that the use of points was a higher motivator for undergraduate students than graduate students.

Public Q&A systems had integrated the use of points and rankings in a leaderboard as a means to motivate user participation, but the opposite effect appears to occur in an education setting. From the open-ended survey responses, students indicated that they did not like to be measured against other students in particular with the use of the leaderboard. Students did not like the feeling of competition among peers. This was reinforced heavily with the Leaderboard being the most cited feature that members did not like on the system. The leaderboard raised a competition aversion issue among students.

58

Undergraduate and graduate students also indicated that they would like their participation points to be tied to something more tangible. Concerning the use of the profile completion percentage, both graduate and undergraduate students mostly disagree with the fact that it encouraged them to update their profile.



| | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|---|
| cs1530 | 13.33% | 20.00% | 40.00% | 16.67% | 10.00% |
| cs2310 | 8.33% | 25.00% | 41.67% | 25.00% | 0.00% |
| combined | 11.90% | 21.43% | 40.48% | 19.05% | 7.14% |
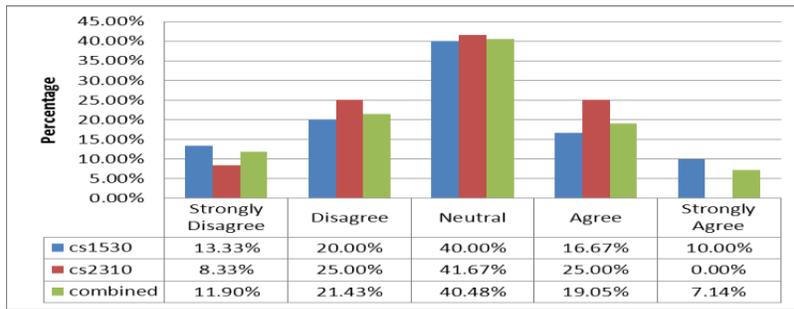
**Figure 11.** Knowing that I can obtain points for asking questions encouraged me to post questions



| | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|---|
| cs1530 | 22.58% | 12.90% | 45.16% | 19.35% | 0.00% |
| cs2310 | 8.33% | 33.33% | 8.33% | 50.00% | 0.00% |
| combined | 18.60% | 18.60% | 34.88% | 27.91% | 0.00% |
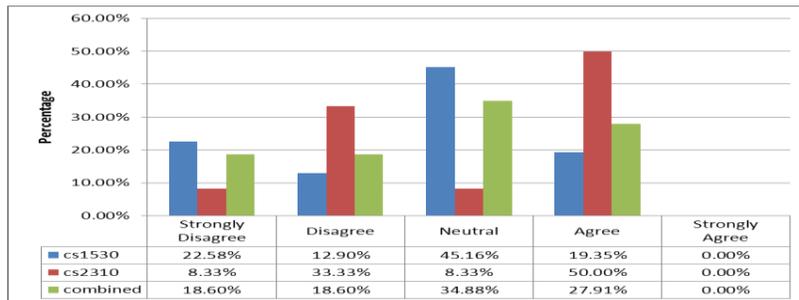
**Figure 12.** Knowing that I can obtain points for answering questions encouraged me to answer questions

### 3.4.2.6 Effect of email notifications on student activity

Students had the option to receive notifications whenever new questions were posted, when the instructor posted an announcement or if their questions were answered.  Figure 13 provides a snapshot on the server activity.  We present activity as either users logging in or students browsing the site to access information.  We can see from the activity log that activity tends to increase as either a question was posted or an announcement was made.  However, in the

absence of such events, activity for the most part tended to decrease. The success of the Q&A systems depends on student participation thus it is important to keep reminding users to come back. For this reason, it was important to ensure students were reminded to come back and check for additional information. From the activity log, we can also see that students and instructors were both creating new content. The second half of the month of November falls during the U.S. Thanksgiving Day holiday, and students begin their Winter Break. Traditionally this is a time of low activity which was corroborated by an absence of new questions being created. However, the instructor made some announcements during this time, which resulted in some students accessing the system.

Thus, we can see how important a role the use of email notifications played in the deployment of MicroProbe. However, we also received some mixed feedback from students. While most students liked being notified of new questions, answers or announcements, some students were inconvenienced by the fact that they could not control the frequency of them.
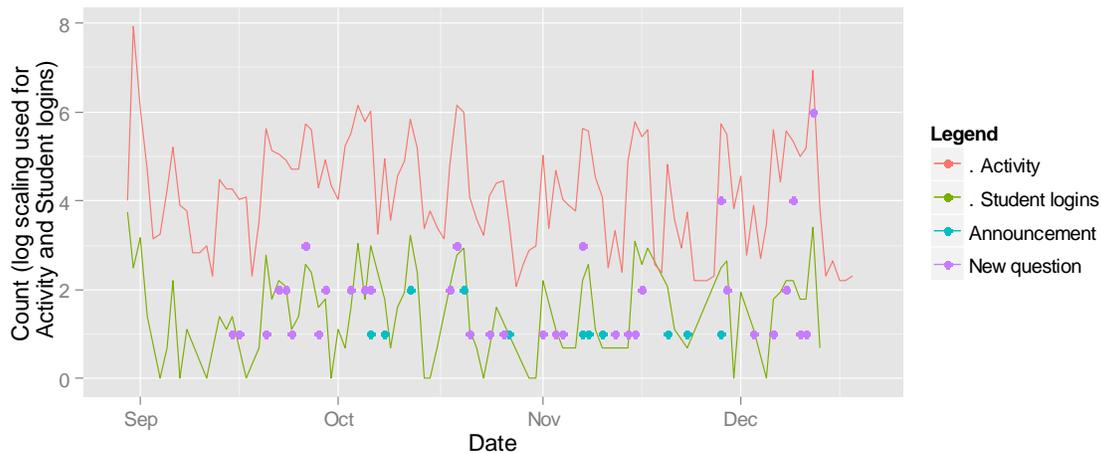


**Figure 13.** Server activity log showing the effect on user behavior as new questions and announcements notifications are received.

60

### 3.4.2.7 Ease of use

The results show students had little trouble adopting the Q&A model. The concept of participation via a Question and Answers system combined with a simple interface showed the greatest impact to users. Sixty two percent (62%) of the students found the system simple to use (Figure 14). Some of the comments stated were: *"The Question-oriented structure, instead of a generic message board makes the purpose of microprobe well defined"* and *"Ability to collaborate with students is \*essential\*. As is the ability to ask/answer questions"*



| | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|---|
| cs1530 | 0.00% | 9.68% | 22.58% | 61.29% | 6.45% |
| cs2310 | 8.33% | 8.33% | 8.33% | 66.67% | 8.33% |
| combined | 2.33% | 9.30% | 18.60% | 62.79% | 6.98% |
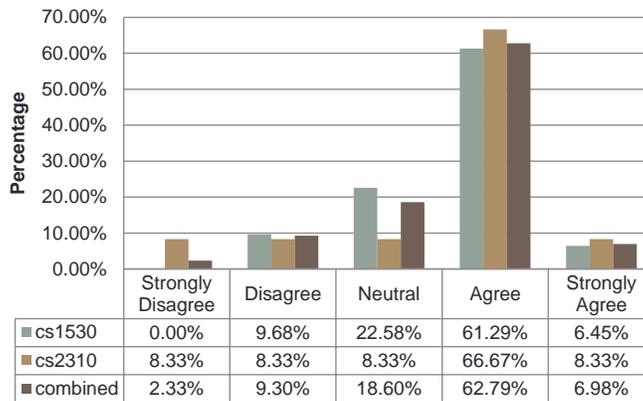
**Figure 14.** Responses to "I found the MicroProbe system to be simple and easy to use"

### 3.4.2.8 New Feature Suggestions

Students identified some usability items and gave several new feature suggestions. The majority of students would like to be able to categorize questions into topics for easy access. Students also would like more control over the frequency of the emails they receive. Privacy and anonymity are also important. Some students suggested the ability to post questions anonymously. This feature might in fact help with the fear of asking questions publicly. Also, a small number of students suggested the ability to access the system from a smartphone.

Finally, the majority of students (40%) indicated a preference to use the system in another class.

### 3.4.3 Lessons Learned

This Chapter examined the use of Q&A systems enhanced with gamification concepts in educational settings. We also aimed to understand the motivators and inhibitors to student participation.

From the introduction of gamification concepts we learned that, in contrast to public Q&A systems, altruism is not a major motivator for undergraduate students to contribute knowledge. The use of points as a motivator to answer questions was viewed more positively by undergraduate students. Graduate students on the other hand, did not like the use of points. We could achieve a higher benefit from gamification if we tied rewards to more tangible *prizes* such as homework or course credit. Also, we wanted to recognize students' participation by the use of leaderboard rankings; however its use was not effective and had more of a detrimental effect on student participation. Instead, we should look at other approaches such as using badges which can meet the need to provide recognition while limiting the amount of comparison between students.

A supporting infrastructure alone is not enough to attract and retain student interest. There needs to be activity to build interest for students to come back to the system continuously. We learned the importance of working in close collaboration with instructors and getting them on board early in the planning process. Getting instructors onboard in the project facilitates the initial boarding process. This early engagement with the instructor helped us by first having the

instructor ask students to use the system as part of completing their first homework thus getting students involved very early. Also, having the instructor use the announcement functionality and answer students' questions on the system enabled them to build trust on the system.

Email notifications are another example of proven techniques to help with student engagement. However, care should be taken to give students the ability to control frequency and content for such notifications. Feedback from students indicated that while receiving emails was good, they were at times feeling overwhelmed if too many emails were received.

We also learned about the proper timing to collect student surveys. To increase the chances of getting the most number of participants taking the exit questionnaire, we decided that the best time for this was at the beginning of the last session prior to the final exam. Our belief was that such time will ensure the majority of students came to class given their motivation to know how to prepare for the exam. We also decided to administer the questionnaire at the beginning of the session to limit exhaustion.

## 3.5 CONCLUSIONS

We presented MicroProbe, an improved Q&A system that integrates gamification concepts to assist in fostering motivation among students in both graduate and undergraduate education (Zegarra, Chang, & Wang, 2015). Results from a deployment of MicroProbe in both undergraduate and graduate courses found that MicroProbe was easy to use and both student levels found benefit on the use of similar systems as part of their educational experience. We also found that, with respect to the use of gamification concepts, undergraduate students were

more embracing of the idea than graduate students. Also, the use of leaderboards was not as good a motivating factor for participation as we had expected. Students did not like the use of leaderboards as it gave them the sense of competition. While students found some motivation on the use of points, they wanted to get more tangible rewards such as course or homework credit. Our results also showed that the participation rate, in the form of asking and answering questions, was higher for graduate students than undergraduate students suggesting undergraduate students tended to be more consumers of information while graduate students were more producers.

# 4.0    THE DESIGN OF MICROBROWSER

Following the experience gained in the evaluation of the MicroProbe application, we wanted to focus on facilitating peer-learning by scaffolding the knowledge generation and discovery process.

This chapter discusses the design goals of MicroBrowser and presents the several components that make up the application. The utility of MicroBrowser is presented in Chapter 5, where we present a full user study showing that MicroBrowser has a substantial impact in facilitating knowledge discovery and reuse among students.

## 4.1    MICROBROWSER DESIGN GOALS

MicroBrowser was redesigned from the ground up with the essential goals of facilitating the knowledge discovery in large knowledge bases and bootstrapping the knowledge generation and knowledge reuse between students. Thus, the MicroBrowser project started with the following design goals in mind:

a. To introduce the concept of design patterns as facilitators for knowledge discovery and generation.

b. To provide an intuitive and useful mechanism to facilitate the creation of new design patterns as insightful reusable knowledge nuggets.

c. To facilitate exploration of large knowledge bases using an interactive knowledge visualization graph proving visual representations of discussions and their connections to other discussions and associated patterns.

### 4.1.1 Addressing faculty scalability

The faculty scalability problem arises when the number of students enrolled in a course becomes so large that it becomes infeasible for a faculty member to be fully engaged in the learning process of every student. In the case of PBL learning via Capstone Projects, such limited involvement can be detrimental to the success of the project (Hung et al., 2012). Peer-learning aims to address this limitation by allowing students to help each other to solve problems or answer questions. By shifting the responsibility of providing feedback from the faculty to the peers, the scalability of the faculty participation is addressed. However, the proper supporting infrastructure is needed to facilitate peer-learning and interaction. The use of discussion forums and Q&A system serve as this infrastructure as it enabled the exchange of messages between peers asynchronously. The structure of the messages enables the description of the problem and the provision of the feedback in the form of an answer or a comment, which could also help to reduce the feedback loop as there are potentially more knowledge providers.

MB system adopts the design properties of Q&A systems in order to provide the supporting infrastructure to enable the exchange of discussions and feedback generation required for peer-to-peer learning. In addition, MB differentiates from traditional Q&A systems by the

integration of design patterns as reusable knowledge nuggets, which allow for the exchange of best practices to solve common recurring problems.

## 4.2 MICROBROWSER USER INTERFACE

It is said that a picture is worth a thousand words. Through visualization techniques, students can use their visual abilities to discover relationships, structures or patterns in discussion threads. For example, when teaching algorithms visually, one feature that was found effective was giving students a high-degree of interactivity (Saraiya, Shaffer, McCrickard, & North, 2004)(Schweitzer & Brown, 2007). MicroBrowser followed the principles of visual data exploration of "overview-zoom-filter then details on demand" and "relate" (Shneiderman, 1996). The anticipated exploration behavior has four steps. First, we provide an overview graph containing all the discussions and design patterns available. The overview graph was complemented with an adjoining textual view with entries corresponding to each data point in the graph. Second, a student can zoom in the overview graph by identifying interesting discussions based on visual attributes or by panning the graph. Third, the student can eliminate unwanted discussions nodes by applying keyword and timeline filtering criteria. We facilitate keeping the focus of students by quickly updating both the visual and textual displays. By clicking on a discussion or design pattern, the student is presented with further details to continue the exploration. And fourth, we facilitated the exploration of related discussions by displaying relationships between discussions and design patterns as connections between nodes.

The user interface of MicroBrowser (Figure 15) is composed of three main components: a knowledge graph view, a discussion listing view and a knowledge timeline. The knowledge graph view provided the overview; the discussion listing view consisted of a textual representation of the nodes and the knowledge timeline allowed students to filter by date information. Following a Details on demand visualization strategy, additional details about a discussion were displayed using the Details discussion view which followed the layout structure of the main window to a great extent to reduce mental workload.
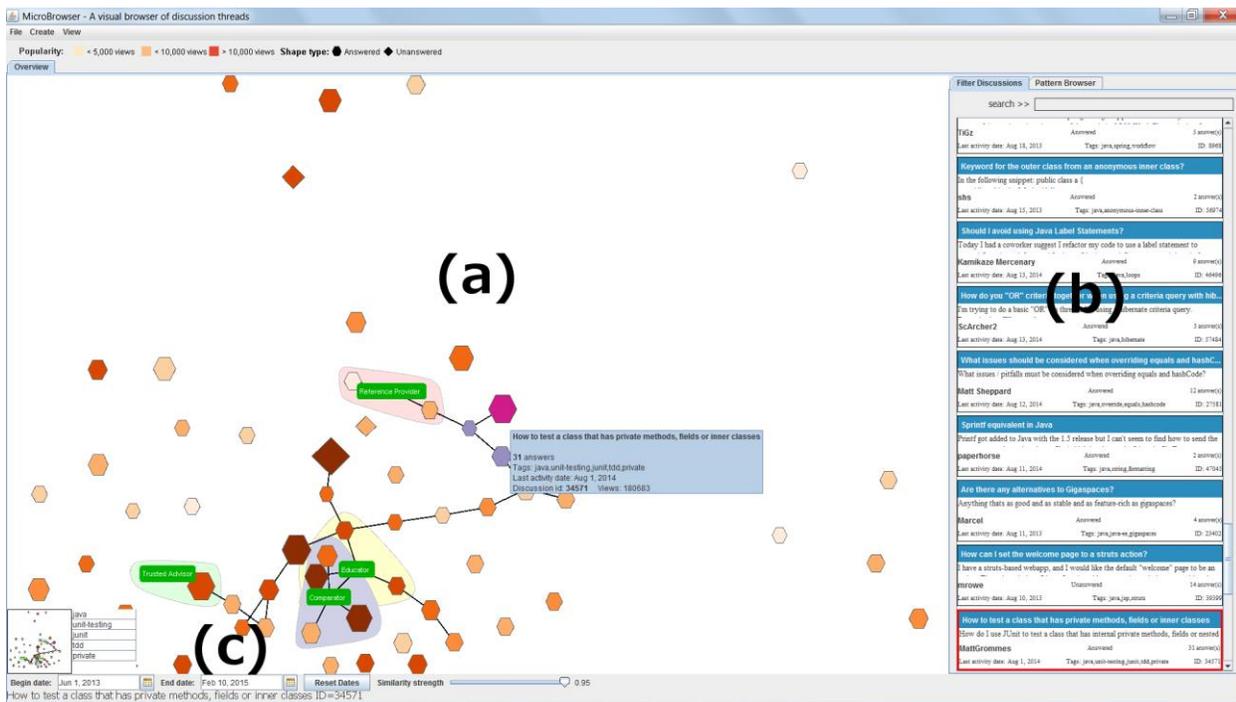


**Figure 15.** MicroBrowser Main User Interface. The main user interface is composed of three main components: A student gains insight over all discussions available from the knowledge graph view (a). Distance between discussions was encoded by topic similarity. In addition to their visual representation, the discussion listing view (b) provided a textual representation of the discussion. The knowledge timeline (c) allowed the student to filter unwanted discussions based on date information.

### 4.2.1    Knowledge Graph View

A visual representation—such as a graph or photograph—is considerably easier to use and to convey more information than a textual description or a spoken report (Shneiderman, 1996). One of the challenges we wanted to overcome was how to facilitate the browsing of a large number of discussion threads within a restricted space while at the same time provide a high-level view of the behavior and trend in discussions as well as their attributes.   Shneiderman suggested the use of an overview strategy in information visualization interfaces to facilitate gaining an understanding of entire collections (Shneiderman, 1996).  Wanner et al. (Wanner et al., 2011)  and Engdahl et al. (Engdahl et al., 2005) followed the overview strategy for displaying entire discussion forum data sets to enable users to gain insights over the behavior and trends in their large collections.  Likewise, MicroBrowser provides the knowledge graph view as a means to implement the overview strategy to gain insights over the entire collection of discussion threads and design patterns as well as their relationships.  By using this approach, students gained a better understanding about the entire knowledge base while facilitating exploratory tasks such as zooming into discussion threads of interest and identifying connections between nodes.  The knowledge overview graph included both discussion threads and design pattern nodes.  The use of direct-manipulation interfaces vindicates the success of the power of using computers in a more visual manner.   Each of the discussion and design patterns nodes in the graph could be manipulated directly– each discussion and design pattern was displayed as a node in the graph that the student could interact with by moving it around the graph, hovering to get summary information and double-clicking it to open up its details.

**4.2.1.1 Discovering discussion state information**

When students search for questions in Q&A systems, they are more interested on locating those that have been answered. At the same time, students looking to contribute their knowledge would focus on identifying unanswered questions that they can supply an answer for. Thus, for Q&A systems such as MicroBrowser, it is important to know whether a particular discussion has been answered or not. Traditional web-based Q&A systems normally provide users with the ability to filter out questions that were answered. In MicroBrowser, we wanted to provide a similar functionality; however, in contrast to traditional Q&A systems, we wanted to allow students to quickly distinguish unanswered discussions by looking at the overview graph. Shape has been used in prior work to encode different information about documents. GistIcons (DeCamp, Frid-Jimenez, Guiness, & Roy, 2005) used the shape to represent the semantic profile of a single document. iBlogVis (Indratmo & Gutwin, 2008) looked at blog entries and used different shapes to denote entries, comments and number of words. MicroBrowser differs from prior work in that we used the shape to represent the state of discussion threads. The shape of a discussion node (Figure 16) indicated whether a discussion was answered or not, with answered ones displayed as hexagons and unanswered ones displayed as diamonds.
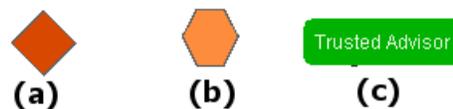


**Figure 16.** Discussion threads without accepted answers were displayed as diamonds (a) while the ones with accepted answers were displayed as hexagons (b). Design patterns were displayed as "label" nodes (c)

70

When displaying design patterns, including the pattern name is important as names make it easier to identify the pattern. Thus, we used a different shape group to differentiate design patterns that contained the design pattern name as part of the node itself. Additionally, to highlight associations to design patterns, we built a convex outline covering those discussion nodes belonging to the design pattern and centered the design pattern node among them.

### 4.2.1.2 Discovering discussions based on their popularity

Users tend to give more attention to discussions that have high popularity than to those with lower popularity. Discussion popularity can provide information about its importance or perceived value to other students. De Choudhury et al. (De Choudhury, Sundaram, John, & Seligmann, 2009) conducted a study to understand communication activity in social networking sites and found that users tend to participate in conversations that have interesting themes and are commented on by familiar users with high social impact. However, multiple other factors could contribute to the popularity of a discussion: author, novelty, hits, author's name, etc. Several methods have been attempted to quantify popularity. Kim et al. (S. Kim & Kim, 2012) measured popularity by a combination of page hits and the time spent reading long discussion blog entries. Carmel et al. (Carmel, Roitman, & Yom-Tov, 2012) focused on measuring the novelty of user-generated content, that is, how popular were contents recently created. Huang et al. (Huang et al., 2014) measured popularity as the number of net votes a discussion received while in DifVis (B. Kim, Johnson, & Baker, 2006), the authors defined popularity as the number of postings per day received by a discussion thread. In MicroBrowser, we use the "number of hits" metric used by Kim et al. (S. Kim & Kim, 2012) and defined the popularity of a discussion as the number of views or hits. This metric is simple and direct since it reflects a student's

attention to a particular discussion thread. We argue that if students tend to access a discussion or question more frequently, they find the question important. Similar to prior work, we allow students to discover discussions by their popularity but extend this capability to discussion threads in Q&A systems.

In addition to being able to measure popularity, it is important to use appropriate means to represent popularity. Color coding has been proven to be an effective means to code quantities in spatial displays(Spence & Kutlesa, 1999). Wanner et al. (Wanner et al., 2011) used 3 equi-depth color bins to represent attributes of a discussion such as the number of hits received. In (Leuski & Allan, 2000), different shades of a color to represent relevance of a document. In MicroBrowser, we encoded popularity using a gradient color scheme with darker hues denoting more popular items.

### 4.2.1.3 Finding relationships between discussions and design patterns

The knowledge graph in MicroBrowser encodes the relationship between discussions and design patterns by using spatial distance and connections. The distance between two discussion nodes in the knowledge graph encodes how similar two discussions are based on their topic similarity. The similarity between two discussions threads was based on the Topic Model generated from the text corpus of the knowledge base and is further described in Section 4.4.2. A connection between a discussion node and a pattern node indicates that a discussion node is an instance of the pattern. A connection between two discussion nodes indicates that the two discussions are related to each other. A challenge we faced with the number of connections between nodes was that it was cluttering the drawing area making it unreadable. We addressed this by using progressive disclosure (Chuang, Ramage, Manning, & Heer, 2012) and added a similarity slider

to allow students to control the visibility of the nodes based on their similarity strength. Details on how similarity strength is calculated are discussed in Section 4.5.1. Also, we used the similarity strength value assigned to each connection or edge between two discussions to determine the color intensity of the similarity with more similar items getting darker hues.

### 4.2.1.4 Preventing discussion nodes overlap

Dimension reduction techniques such as Multi-Dimensional Scaling (MDS) seek to facilitate the creation of visual representations of multi-dimensional data by reducing the number of dimensions to smaller (1D, 2D or 3D) visualization spaces. In MicroBrowser, we used MDS to reduce the similarity between discussion threads to a 2-dimensional space. While such techniques are useful, they present an overlap problem when trying to visualize highly related items in a reduced dimension. Force-Directed Placement is a technique commonly used in graph visualization that uses simulated physical forces to prevent node overlap (Eades, 1984). As forces in the graph become stabilized, the overlap between nodes is reduced. Prefuse made available an implementation of force-directed layout based on the algorithm described in (Barnes & Hut, 1986). We configured the force-directed layout using the parameters of gravitational constant -0.8f, minimum distance 360f, and default theta for node repel and a spring force coefficient of 1e-5f and default length of 5 for edge spring behavior. Because force-directed layout used connections between nodes to apply spring forces, we faced a challenge as some nodes did not have any connections to other nodes. This situation was present, for example, when filtering criteria was applied to remove unwanted discussions. To prevent nodes that were not connected being pushed too far away by the algorithm and to avoid losing the spatial relationship encoding, we only applied the forced layout during the first 2 seconds after

the knowledge graph contents were loaded. While spatial information was lost, we felt it was acceptable given that students would not be presented with overlapping discussions that obscured information.

## 4.2.2   Discussion Listing View

The canonical browse and search technique used in the Web has been mostly text-based using interaction techniques such as scrolling and mouse clicks (Sebrechts & Cugini, 1999). Text-based browsing schemes provide a simple and effective means for displaying textual information that cannot be fully represented using visualizations in a graph node. The discussion listing view (Figure 17) provides a text-based browsing scheme to allow students, especially those more familiar with Q&A systems, to access more information about the discussions that would have been difficult to include in the graph while at the same time reducing the learning curve found by only providing the visualization. For each discussion entry we displayed the full question title, up to 200 characters of the question body, author's display name, whether it was answered or unanswered, the number of answers, the last activity date, the list of tags and their unique Id. Supported interaction techniques included scrolling, mouse over and double-clicking. Double clicking on a discussion entry opened its detailed view.
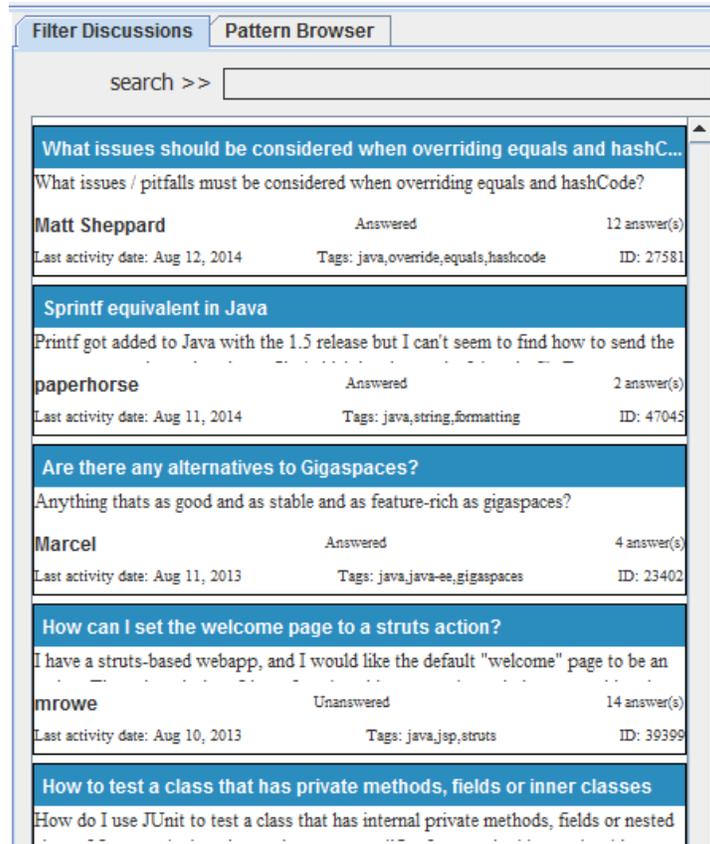
**Figure 17.** Discussion listing view supporting text-based browsing scheme. Each entry shows question text, title, author, state, number of answers, last activity, and tags. Students can use the search input to filter discussions based on keywords.

### 4.2.2.1 Filtering discussions by keywords

Browsing through the discussions might not be feasible for large data sets. To assist students in removing unwanted discussions, we provided a keyword-based search feature. Such features require minimal effort from students, requiring the input of a set of words. To filter discussions, students simply entered one or more keywords and pressed the enter key. We made use of the

Apache Lucene[5] text search engine library and built the search index by a combination of the question title, body and the question tags. The keyword-based search was case insensitive.

### 4.2.2.2 Synchronized navigation

Since we provided two different browsing methods (knowledge graph and text-based discussion listing), it was important to ensure proper integration between the two to ensure a seamless user experience when browsing discussions and design patterns. To support this integration, we enabled bi-directional navigation between the knowledge graph and the listing view. With this technique, as students moved the mouse over a node in the graph, the corresponding entry in the discussion listing view was brought into focus. Likewise, if a student was more comfortable browsing discussions in textual form, as they moved the mouse over the discussions the corresponding node in the graph became highlighted.

### 4.2.3   Knowledge Timeline

Asynchronous communication between users of public online knowledge bases tends to occur at times that are convenient for them and usually overlooks temporal norms with users normally disregarding time cues (Gibbs et al., 2006). Contrary to asynchronous communication on public knowledge bases, we believe that in an educational setting, temporal information plays a more important role. Seaton et al. (Seaton et al., 2014) analyzed student behavior in a Massive Open Online Course (MOOC) and found that discussion forum activity in MOOCs increased over time and peaked prior to midterms and finals. In separate study performed by Warren et al, the

---

[5] http://lucene.apache.org/core/

authors found that activity on class forums for an online course were very activity 4-12 hours before deadlines (Warren, Rixner, Greiner, & Wong, 2014).   These results suggest that students not only use forums in an ad-hoc manner to find help with course problems, but there is a temporal component associated to their use because students consider discussions occurring in forums during particular time frames as sources  of knowledge that can contribute to their preparation.   Thus, we can see that by facilitating access to the temporal properties of discussions, we are contributing to their learning.  Giguet et al. (Giguet & Lucas, 2009) and Gibbs et al. (Gibbs et al., 2006) used temporal information to monitor collaboration between users in forum activity.  In MicroBrowser, we were more interested on the discovery process as opposed to understanding collaboration between students.  The knowledge timeline allows students to filter discussions by specifying starting and ending dates.  For example, let's assume a student wants to find activity since the midterm.  A traditional Q&A system facilitates finding the most recent discussions but finding activity occurring within particular timeframes is not easily done without having to scroll through all discussions sorted by date.   Using MicroBrowser, the student can simply set the starting date to the start of the midterm, and the knowledge base will be quickly set to only those activities since that date.  Date selections are inclusive.  Using the timeline in combination with the keyword filters, students can further narrow the available discussions looking for more specific discussions.

## 4.3    DISCUSSION DETAILS VIEW

The discussion details view provides the "details-on-demand" visualization feature in our design. While the knowledge graph view provides an overview of all discussions, we wanted to give students the ability to drill down to either a discussion or a design pattern to obtain more detailed information. To maintain consistency in the interaction behavior from the overview we kept the layout of the details page similar but with some modifications. For our discussions, we wanted to represent an organization of answers, related discussions and associated design patterns. Cai et al. (Cai & Sheth, 2003) noted that using a tree layout provides a simple way to represent perceptual organization. In MicroBrowser, we similarly built a perceptual organization and hierarchy for answers, related discussions and design patterns (Figure 18) using the tree layout. The discussion detail view also included the color and size encodings from the knowledge graph view.
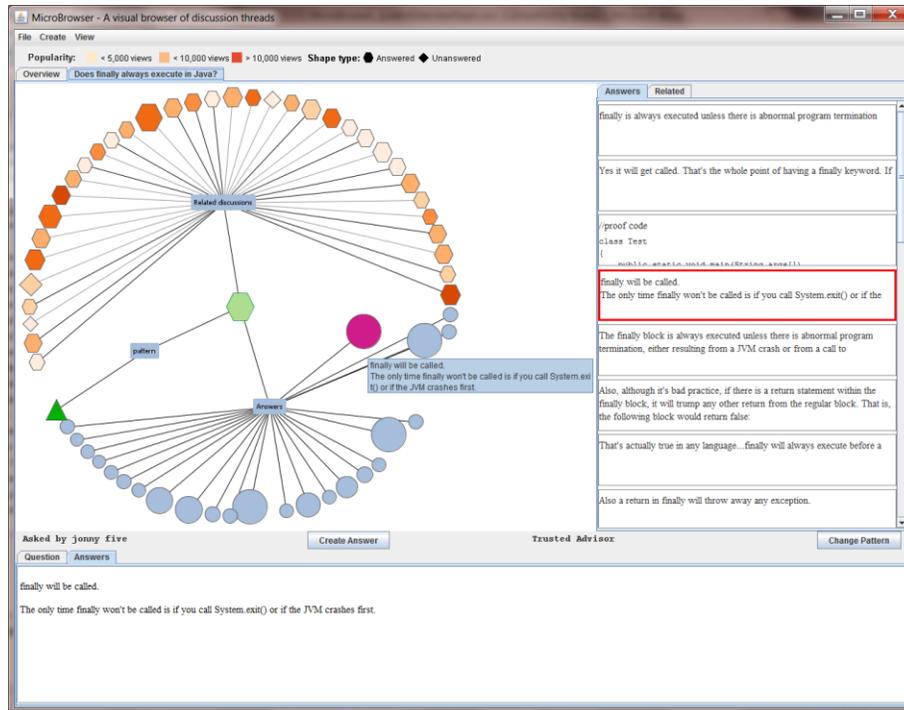
**Figure 18.** Discussion details view providing question details, answers, related discussions and design patterns. In this figure, a student has selected an answer. The selected answer is brought into focus on the answers list to the right and its content is displayed in the Answers tab below.

In addition, similar to the knowledge graph view we presented textual representations of the nodes in the graph. The contents of the question and its answers could have a large amount of information. To display this content, we added a tabbed layout at the bottom where we displayed the contents of either the question or the answer. To load text content, a student only needed to select the desired answer from the graph or from its textual equivalent in the text browsing view. From the details view, a student could also create a new answer and associate a design pattern to the discussion.

79

### 4.3.1 Finding the accepted answer and other important answers

Being able to identify the accepted answer for a discussion is a critical step in knowledge discovery. MicroBrowser facilitates the discovery of the accepted answer by first displaying all answers under the *answers* meta-node and then using color to differentiate the accepted answer from other answers. Using this approach, when students first open the details of the discussion, they can quickly identify the accepted answer from the other answers by their color. Sometimes, the accepted answer might not provide enough information to other knowledge seekers. Thus, we also made use of the size of the answer nodes to encode the number of up-votes. By this method, if the accepted answer does not provide enough information for a related question or enough explanation for a question, a student could quickly refer to other important answers by identifying their shape size.

### 4.4 INTEGRATING DESIGN PATTERNS IN KNOWLEDGE DISCOVERY AND GENERATION

While Q&A systems are technically advanced, they do not address key pedagogical goals (Duffy, Dueber, & Hawley, 1998) and thus could be improved. Design patterns are a shared language used to communicate proven solutions to recurring problems. In its basic form, a design pattern is composed of a name, problem, solution and a context. The name identifies the design pattern while the problem describes a recurring problem that occurs within the design pattern context. Problems are described in abstract form to facilitate their applicability to

different domains and situations. The solution to the problem is a recommended approach to solving the problem. If we look at questions as being problems that occur in the context of an educational setting, then we can build an analogy that answers to the questions represent their solutions. As with FAQ lists, some questions in Q&A systems tend to be visited or accessed more frequently than others (Wang et al., 2013). While there could be several explanations for this behavior such as novelty (Carmel et al., 2012) or author's reputation (Coetzee et al., 2014), the popularity of a question can be influenced by other students having either the same or a similar question. At the same time, Mamykina et al. (Mamykina et al., 2011) and Huang et al. (Huang et al., 2014) found that user activity follows a power law, with a few subset of users answering most questions. Those power users, over time, are able to identify trends in questions being asked which might look similar to other questions but with different details. We can then see that there is a degree of recurrence of repeated questions being asked with different details but under the same context. Thus, by associating discussions to design patterns we indicate that the design pattern problem is an abstract representation of the discussion question. With context being the educational setting, we can then refer to the solution of the design pattern as a possible solution to the original discussion question. Different from merely tagging a discussion, by associating a design pattern to a discussion we provide a comprehensive and holistic picture of the strengths of the given solution for the question. Our goal was to bootstrap novice users with the ability to build associations between questions by looking at them in a more abstract way, something that power users have gained over time. In pattern oriented instruction, Muller proposed the use of algorithm patterns as an approach to enhance analogical reasoning in students to develop problem solving skills (Muller, 2005). In MicroBrowser, design patterns facilitate peer-learning and knowledge sharing by the support of three important properties:

knowledge discovery, knowledge creation and knowledge reuse. Knowledge discovery is facilitated by the ability to find related questions or discussions based on context and problem type similarities. Knowledge creation is facilitated by helping students to create new design patterns as well as by the ability for students to create relationships between discussions and design patterns. Finally, knowledge reuse is facilitated by allowing students to reuse the solutions from design patterns to provide answers to similar questions. Thus, as shown in Figure 19, through the use of design patterns and the supporting infrastructure, we define a loop that not only contributes to the generation of new knowledge but it also enables peer-learning and knowledge sharing.
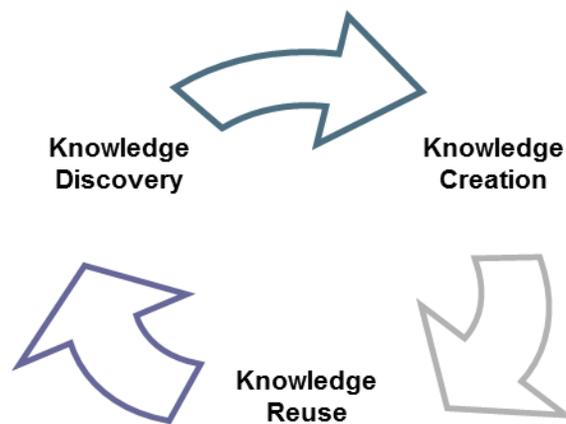


**Figure 19.** Design patterns facilitate peer learning and knowledge sharing.

### 4.4.1.1 A usage scenario on using design patterns

Students used design patterns for knowledge discovery by locating discussions that were related to that pattern. Using the knowledge graph, students had access to the connections between discussions and design patterns. In addition, looking at the details of a design pattern, a student

could see related discussions. We can interpret related discussions as being instantiations or examples of the design patterns.

Design patterns facilitate knowledge reuse by being able to follow the solution recommended by a design pattern to answer a question that is described by the design pattern problem. For example, assume that a student was working on a research project that investigated subjects' reactions to images presented by a computer. The advisor of the student told him that instead of writing a computer program to pick the images, to ask a lab mate to pick images instead. The students sees the benefits of this approach and creates a design pattern called *Wizard of Oz* to describe a problem where a user wants to simulate the behavior of a system and the solution where a human is asked to take the role of the computer. Now, if another student is working on a different research project and is looking for recommendations on how to write a program to perform some computer behavior quickly, by referring to the *Wizard of Oz* design pattern, the student can follow the solution and instead of writing a computer program, the student can ask a lab mate to do the role.

As another example, let us assume a student posts a question asking for a comparison between two concepts. The answerer might recognize the type of question, and looking at the pattern browser, the answer reviews the Comparator pattern. Referencing the solution from the Comparator pattern, the answerer submits an answer. Once the answerer submits the question, the asker will review it and consume the new knowledge. If we establish a relationship back to the pattern composition, we have as our context a student having a question while attending a course and a problem of a question of comparing between concepts. The solution would consist of how to structure the answer back to the student in a form that not only answers the question but does it in a manner that the other students would find it helpful as well.

**4.4.1.2 Discovering patterns**

Deng et al. (Deng et al., 2005) suggested that tools that manage pattern collections should be able store and organize the patterns for easy access and exploration. Several approaches have been used to display the list of available design patterns to users such as visual hierarchies (Díaz, Aedo, & Rosson, 2008) and pattern lists (Lin & Landay, 2008). The Pattern Browser view (Figure 20) provides a listing of available design patterns enabling students to explore them. Each design pattern entry displays the name of the pattern and its problem or description. The Pattern Browser is available from the main navigation window from the discussion listing view.
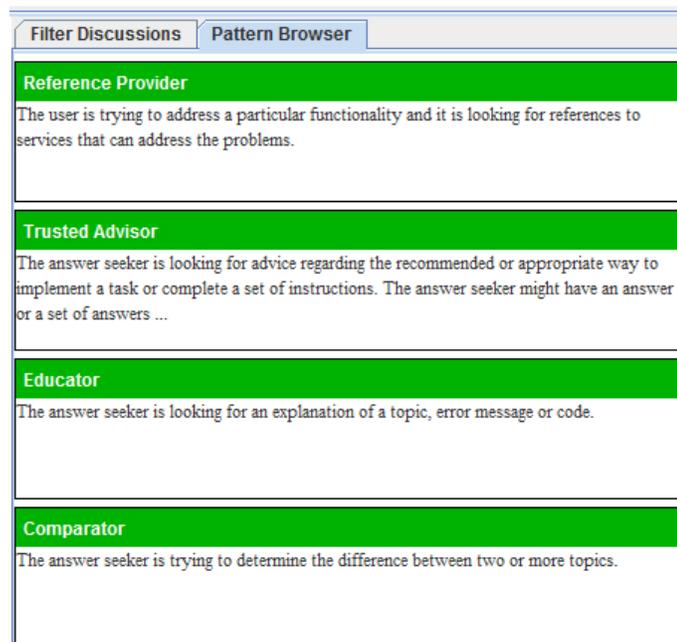


**Figure 20.** MicroBrowser's Pattern Browser displays all available patterns for use. It was accessible from the overview window. Each pattern entry displayed the Pattern Name and Patter Description. Double clicking an entry opened a detailed view.

Design patterns can also be explored from the knowledge graph view as shown in Figure 21. Students can identify design patterns by their green labels displaying the design pattern name. To highlight associations between discussion nodes and design patterns in the knowledge graph, we used two approaches. First, we displayed a connection between the two nodes in the graph providing a visual cue of the association. Second, to show association of a group of nodes to a design pattern, we displayed a convex circle spanning all discussion nodes associated to the same design pattern.
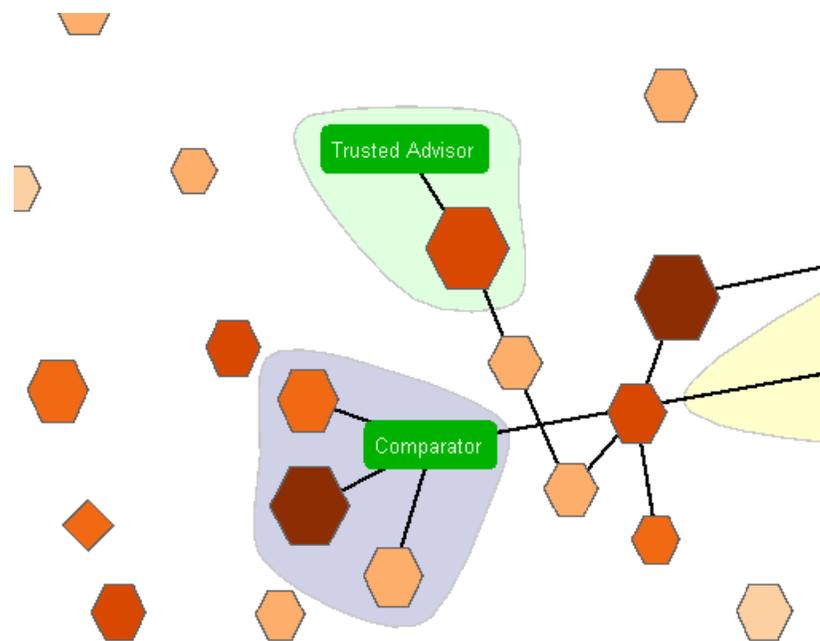


**Figure 21.** Visualizing design patterns and their connections

### 4.4.1.3 Associating design patterns to discussions

The pattern selection dialog, Figure 22, presents students with a list of available patterns they can select from. Selecting a pattern displays its description and the solution. The description helps

the student identify the appropriate question type being evaluated. The solution provides the student with a recommendation for how to answer the type of question or problem.

Once a pattern is selected, the discussion becomes an instance of that pattern. If students need to find samples of how similar questions were answered, they can look at other instances of discussions associated to the selected pattern. It is possible that students are not able to identify any appropriate design patterns from the list. For this scenario, we included an option to "create new pattern …" which enables the students to define a new pattern.
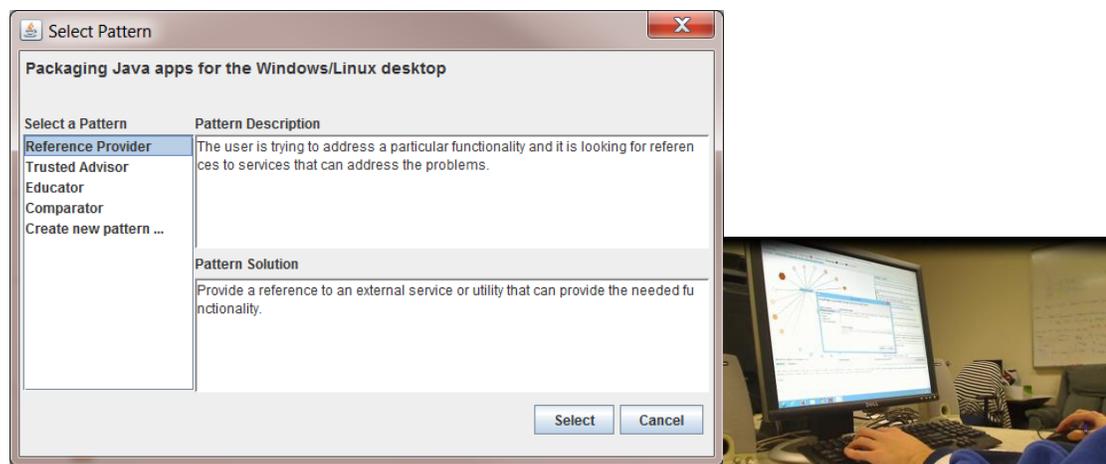


**Figure 22.** Pattern selection dialog allows students to select existing or create new design patterns. In the next picture, we see a student using the Pattern Selection dialog to associate a design pattern to a discussion.

### 4.4.2    Creating new patterns

Knowledge evolves slowly over time as new facts are acquired and new lessons learned (Zegarra et al., 2010). Similarly, design patterns need to be able to evolve and grow over time to capture these new approaches and techniques to address new problems. Students contribute in this process by the identification and creation of new design patterns. MicroBrowser facilitates the creation of design patterns in two different ways. Students can directly invoke the Create New

Design Pattern dialog from the main menu bar. The dialog is shown in Figure 23. Alternatively, from the Pattern Selection dialog (Figure 22), if students cannot find an existing design pattern, they can select the *Create new pattern ...* option which will open up the create design pattern dialog.

To create a new design pattern, the student must provide a pattern name, a pattern description and a pattern solution. Once this information is supplied, the student clicks on the *Create Pattern* button.
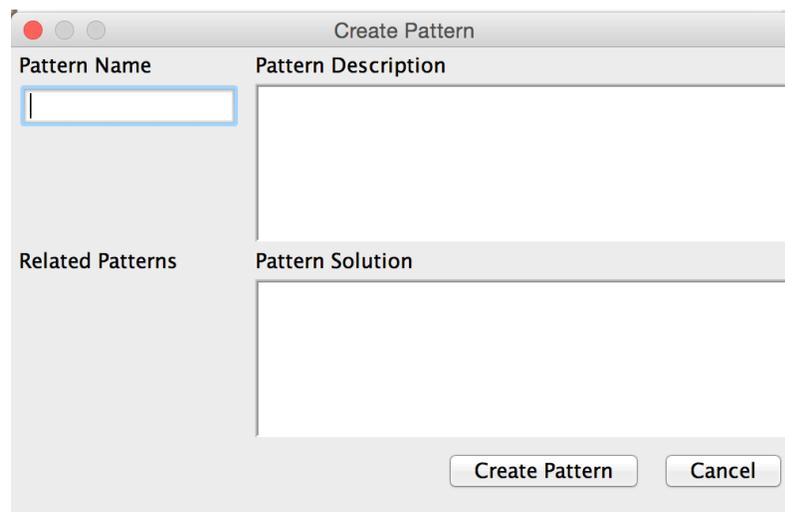


**Figure 23.** Create New Pattern dialog.

Once a new design pattern is created, it is automatically added to the knowledge graph, knowledge listing and pattern browsers view. If the design pattern was created from the Pattern Selection dialog and the student associates it to the discussion, then the new design pattern is automatically associated to the discussion.

## 4.5   DATA PROCESSING AND GRAPH GENERATION

### 4.5.1   Computing similarities from Topic Models

Topic modeling facilitates the analysis of large volumes of unstructured text. A topic model applies statistical methods to help identify topics that occur in collections of documents. It is based on the basic premise that given a document which contains the presence of words commonly associated to a topic we can then infer that the document is about that topic. A challenge we had in MicroBrowser was to define relationships between the different discussions. Fortuna et al. (Fortuna et al., 2005) provided a general approach to visualize text document collections using topic models extracted from corpus of text documents. In MicroBrowser, we followed a similar approach and used the information from topic models to quantify similarities between discussions. For our text corpus, we created a text document for each discussion by combining the question title, question text, tags and the answers' text. Stop words were removed from the documents. We used the Mallet (McCallum, 2002) topic modeling toolkit to perform topic analysis and build our topic model. The initial model parameters were set to the default values except for the maximum number of topics which was set to 100 and the maximum iteration was set to 2000. Using the topic model, we created a dissimilarity matrix between each discussion. We used (1) to compute the dissimilarities between two discussions to populate our dissimilarity matrix.

$$dissimilarity(i,j) = 1 - similarity(i,j), \quad where \qquad (1)$$
$$similarity(i,j) = \frac{similar\_topics[i][j]}{count\_topics[i]}$$

$similar\_topics[i][j] =$ Number of topics that are similar between discussions $i$ and $j$

After computing the dissimilarities, we followed the approach from Trampuš et al. (Trampuš & Grobelnik, 2010) and applied Multi-Dimensional Scaling (MDS) to the dissimilarity matrix to reduce the discussions to a 2-dimension space for plotting. We stored the resulting similarity value between discussions separately to be used when retrieving related discussions.

### 4.5.2   Capturing knowledge using GraphML

We used the GraphML[6] format to describe the structural properties of the knowledge base with nodes representing discussions and edges representing relationship. For each discussion, we created a connection to each of its top 20 most similar discussions. For the case of design patterns, we built a connection to each associated discussion. For each edge created between two nodes, we stored the calculated similarity value, except for design pattern edges where we set the value to 1.

**Table 2.** GraphML schema definition for nodes and edges.

| Schema Element | Attribute name | Attribute type |
|---|---|---|
| Node | id | int |
| | type | int |
| | title | String |
| | body | String |
| | solution | String |
| | x | double |
| | y | double |

[6] http://graphml.graphdrawing.org

| | viewcount | int |
|---|---|---|
| | answercount | int |
| | acceptedanswerid | int |
| | owner | String |
| | lasteditdate | long |
| | lastactivitydate | long |
| | tags | String |
| | score | int |
| | searchfield | String |
| Edge | id | int |
| | type | String |
| | similarity | int |
| | lasteditdate | long |
| | lastactivitydate | long |

## 4.6    MICROBROWSER TECHNICAL IMPLEMENTATION DETAILS

The MicroBrowser application was redesigned to provide a richer user experience than a traditional Web-based application. The new design provided a standalone rich user interface supporting the visualization of discussion threads and design patterns and their relationships. MicroBrowser consisted of four main components: Data Management, Graph Manager, Data Analytics and User Interface. Figure 24 shows the four components.
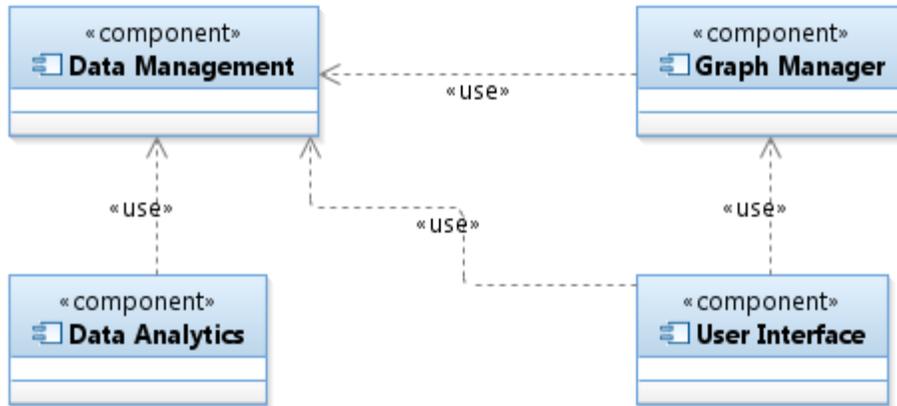
**Figure 24.** High levels components making up the MicroBrowser Project.

The Data Management component contained the scripts needed to extract data from StackExchange as well as SQL scripts to create and manage the database management system. Extracting the data from StackExchange consisted of two steps. Using the script from Listing 1, we extracted the top 500 most recent posts for the tag Java. For each post result we then extracted their answers. We stored the results into our Posts dataset. The fields obtained for each Post dataset entry included: Id, PostTypeId, AcceptedAnswerId, ParentId, CreationDate, Score, ViewCount, Body, OwnerUserId, OwnerDisplayName, LastEditorId, LastEditorDisplayName, LastEditDate, LastActivityDate, Title, Tags, AnswerCount, CommentCount, FavoriteCount, ClosedDate, CommunityOwnedDate.

**Listing 1.** Script used to retrieve the 500 most recent posts for tag 'Java' from DataExchange.

```
declare @tagId int;

select  @tagId = Id
from    Tags
where   TagName = 'java';

     select top 500 p.*
     from posts p join PostTags pt on pt.PostId = p.Id
```

91

```
      where
        p.PostTypeId = 1 and
        pt.TagId = @tagId
        union
  select p1.*
  from posts p1
  where
    p1.ParentId in (
      select top 500 p.Id
      from posts p join PostTags pt on pt.PostId = p.Id
      where
        p.PostTypeId = 1 and
        pt.TagId = @tagId
      )
    and p1.PostTypeId = 2
```

Once we retrieved the Posts dataset, we then proceeded to extract information about users

using the script in Listing 2. To retrieve user information only for those users that had activity in

the Posts dataset, we reused the same query but matched the results against the Users table. For

each result, we extracted the Id, DisplayName and Reputation.

**Listing 2.** Script used to retrieve user information from the 500 most recent posts and their answers.

```
declare @tagId int;

select  @tagId = Id
from    Tags
where   TagName = 'java';

select u.id, u.displayname, u.reputation
 from (
   select top 500 p.*
  from posts p, PostTags pt
    where
        p.Id = pt.PostId and
        p.PostTypeId = 1 and
        pt.TagId = @tagId ) t join users u on t.OwnerUserId = u.Id

union

select u.id, u.displayname, u.reputation
 from (

select p1.*
```

```
  from posts p1
 where
   p1.ParentId in (
     select top 500 p.Id
     from posts p join PostTags pt on pt.PostId = p.Id
     where
       p.PostTypeId = 1 and
       pt.TagId = @tagId
     )
     and p1.PostTypeId = 2 ) t join users u on t.OwnerUserId = u.Id
```

To persist and manage our data we used a relational database. We used the Derby Database Management System to store our data models. The schema used for our data model is captured in Figure 25. The schema consisted of 7 tables: THREADS, USERS, PATTERNS, SIMILARITIES, SVD_THREADS, PATTERNS_2_POSTS and PATTERN_2_PATTERN. The THREADS and USERS table contained information about the discussions and users, respectively. We initialized these tables with the Posts and Users dataset which we obtained from StackExchange. The PATTERNS table captured information about patterns. The SIMILARITIES table captured similarity values between two discussions. We captured results from similarities and Singular Value Decomposition (SVD) calculations into two separate tables to facilitate the graph re-generation process. The SVD_THREADS table captured the X and Y coordinates resulting from SVD calculations.
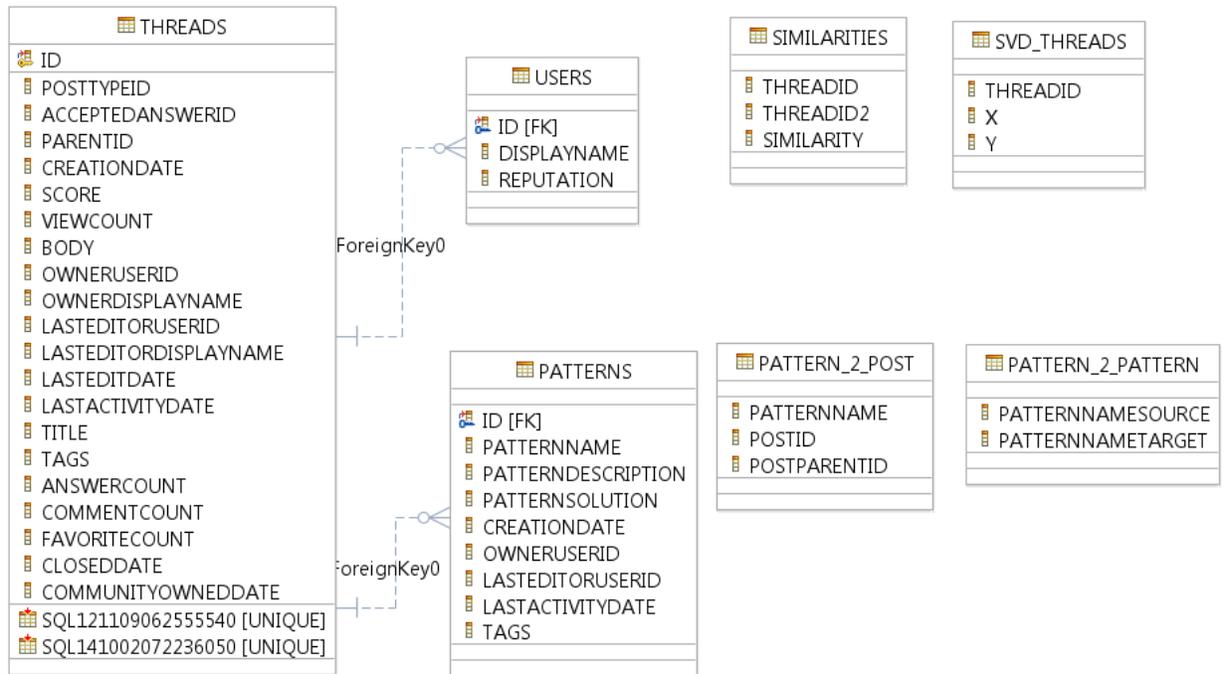
**Figure 25.** MicroBrowser Schema Model Overview

To capture relationships between discussion threads and design patterns, we used two tables PATTERN_2_POST and PATTERN_2_PATTERN. As students associated design patterns to posts, a corresponding entry was added to the PATTER_2_POST table. Design patterns could also have relationships to other patterns and form pattern languages. To capture this relationship, we used the PATTERN_2_PATTERN table.

The Data Analytics component handled the pre-processing of the data. In particular, it handled the generation of the similarities and SVD calculations. This component integrated Mallet Topic Modeling Toolkit (McCallum, 2002) scripts to generate the topic model from the discussion corpus, Java Machine Learning Library (Java-ML)[7] to perform Cosine similarity

---

[7] http://java-ml.sourceforge.net/

calculations and the Multidimensional Scaling for Java[8] (MDSJ) library which provided Multidimensional Scaling (MDS) algorithms used to perform dimensionality reduction for data visualization. Further details about the inner details of this component is discussed in Section 4.4.2.

The User Interface component of MicroBrowser was written in the Java programming language and it is primarily based on the Prefuse (Heer, Card, & Landay, 2005) data visualization toolkit. Figure 26 provides a high-level class diagram for the application. Following the Prefuse project structure, we defined packages to contain control, render, and action classes. We also defined a package for storing utility libraries and finally, the main component contained the main application along with constants and configuration files.

---

[8] http://www.inf.uni-konstanz.de/algo/software/mdsj/
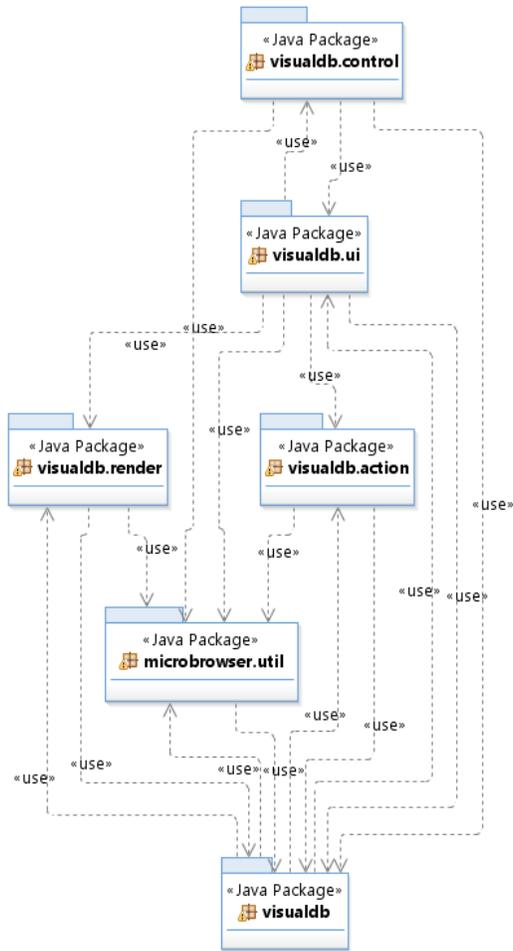
**Figure 26.** MicroBrowser UI component package diagram

To enable different behavior within the visualization, the framework required the use of Action classes. Action classes facilitated the modification of the properties of the elements in the graph. Actions could run infinitely or be triggered by events. Prefuse provided common Action classes such as Search; however, we added custom actions to be used by MicroBrowser as shown in Figure 27. These included Action classes to control the changes on the properties of the discussion nodes, for example, shape color, shape text, node font, node border color, edge stroke color, among others. Each of these actions would change the appearance of the nodes in

response to actions such as hovering over a node, or doing single or double clicks.   To control

the placement of nodes in the graph, we used the PatternLocationAction whose main task was to

center pattern nodes in the center of all member discussions.   Depending on the search criteria or

time line selections users made, we needed to show different nodes in the graph.   We defined

visibility Actions for example, the PatternVisibilityAction and EdgeVisibilityAction which

controlled the visibility of the design patterns and the edges connecting the nodes respectively.

For controlling the visibility of nodes, we made use of Prefuse's PredicateFilters in combination

of visibility actions.  The UpdateListItemsAction class implemented the support needed to keep

the graph and the text browsing schemes synchronized.
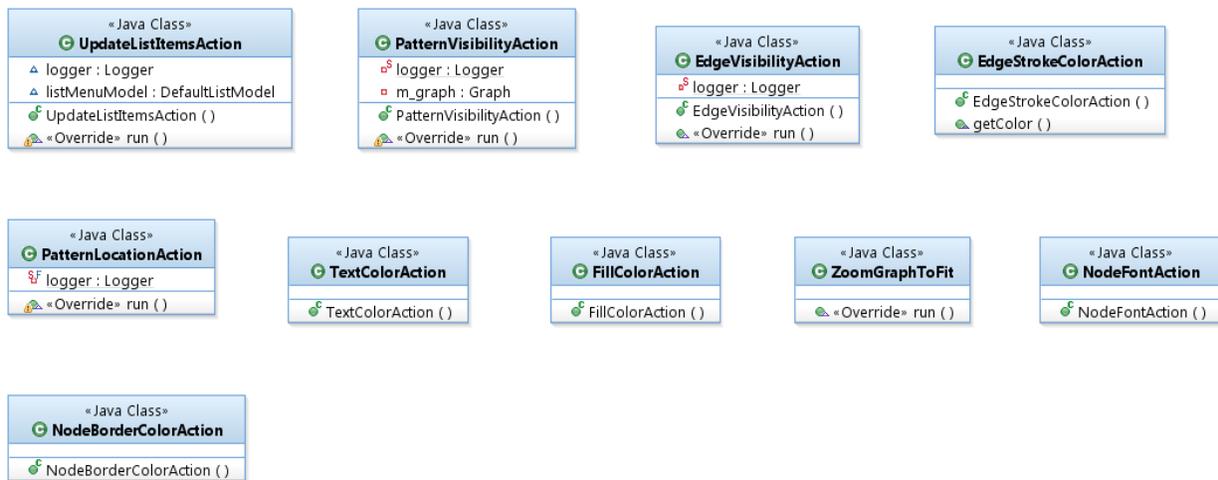


**Figure 27.** Class diagram for Actions package

The appearance of the different nodes in the graph was controlled using Renderer classes.

Using Java Graphics2D classes, Renderers controlled the visual appearance.  Figure 28 presents

the different renderers we defined to control how Discussions, Edges, List Items, Patterns and

Answers were displayed to students on the different views.
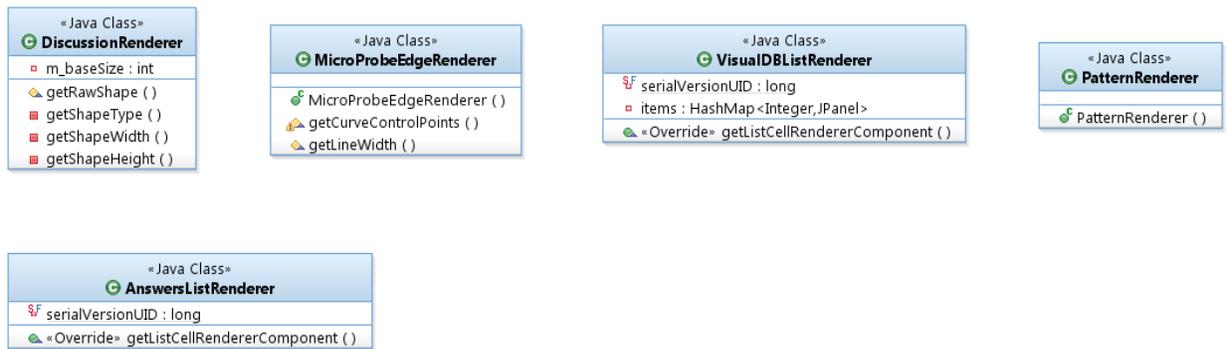
**Figure 28.** Class diagram for Renderer package

MicroBrowser had several dialogs and views that managed the data and provided different perspectives into the data. The first view that students saw when accessing the application was defined by the DiscussionOverviewBySimilarityPane class. This class displayed the overview of all the discussions and associated patterns valid for the keyword and timeline settings. The view provided both the interactive graph and the text browsing scheme enabling students two ways to browse the discussions. Students could also see discussion nodes organized by Patterns. This view was implemented in the DiscussionOverviewByPatternView class. If a student wanted to ask a question, the system instantiated the AskQuestionDialog class. Opening the details of a discussion node loaded the DiscussionDetailsPane class. The PatternsDetailsPane class was a subclass of it and was invoked when the node selected was a Pattern. When a student wanted to submit an answer, the system displayed the SubmitAnswerDialog class. To change the associated pattern, the system displayed the SelectPatternDialog. To edit a Pattern, the system displayed the EditPatternDialog and whenever a student wanted to create a new pattern, the application used the CreatePatternDialog class.
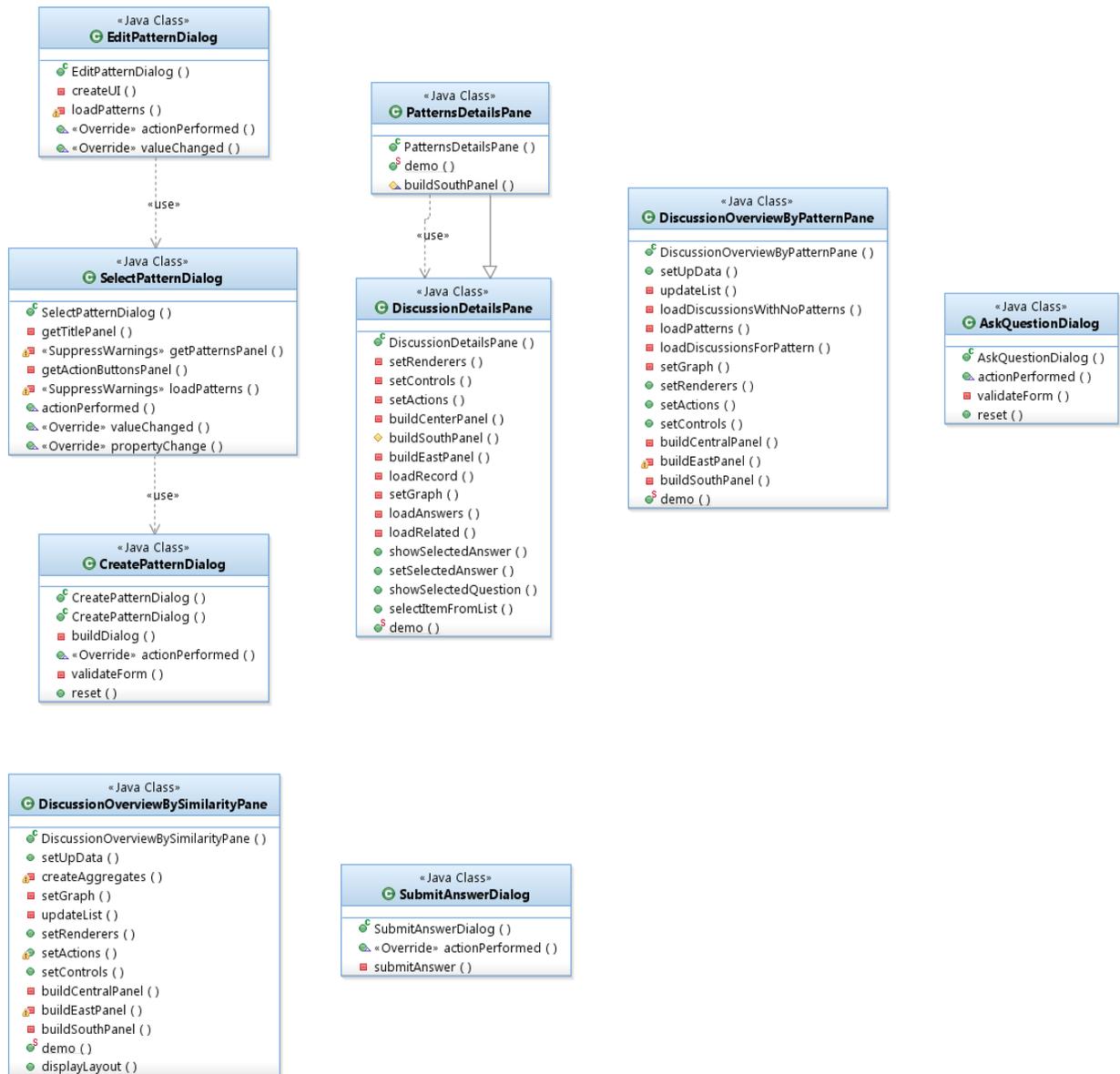
**Figure 29.** Class diagram of UI dialogs and panes

## 4.7 CONCLUSION

In this chapter, we provided a description of our innovative system named MicroBrowser and presented how design patterns are introduced as part of the knowledge discovery and generation process. We also described how our design pattern aware system leverages the features of design patterns and facilitates the use of them by students.

# 5.0     EVALUATING MICROBROWSER WITH STEM STUDENTS

In Chapter 4 we described our proposed innovative system MicroBrowser (MB). After implementation of our system, the next goal was to evaluate it. The primary goal of this chapter is to describe and present the results of a user study performed to evaluate the feasibility and effectiveness of MicroBrowser among students in Computer Science Education (CSE) and Information Sciences Education (ISE) majors. To evaluate the feasibility of our system, we conducted a user study to determine if MicroBrowser:

- Promotes peer learning and knowledge discovery and whether students were able to discover and reuse knowledge efficiently

- Facilitates completion of knowledge discovery and reuse tasks, and if students understood the concept of design patterns

- Improved user experience for browsing discussion threads with its novel design features of color coding, node sizes, keyword and timeline search and discussion overview

- Was easy to use and whether students would consider integrating it into their educational setting

We also compared it against a state-of-the-art Q&A system (Q&A) to assess the effectiveness of MB.

## 5.1    METHODOLOGY

For the user study, we used a within-subject design using two systems: MB and Q&A. We recruited 32 participants among CSE and ISE majors from the Department of Computer Science and School of Information Sciences from the University of Pittsburgh students to participate in our study. Inclusion criteria consisted of being at least 18 years old, have some experience with programming languages and have familiarity using computers.

Each study session lasted 1 hour 30 minutes. Participants received a $10 Amazon.com gift card for their participation. All sessions took place in a lab room at the Department of Computer Science at the University of Pittsburgh.

### 5.1.1    Apparatus

Participants used a Dell Optiplex 745 (Intel Core2 Duo T6300 1.86GHz, 2GM RAM) using Windows 8 and a 19" display at a screen resolution of 1280x1024. MB ran on Eclipse V4.2.2 and Java SE 7u51. As our baseline Q&A system, we configured an instance of the open source, PHP based Question2Answer[9] platform. Participants accessed it using an Internet Explorer V10 web browser.

---

[9] http://www.question2answer.org

### 5.1.2 Data for the User Study

The data for the user study consisted of the 500 most recent discussion threads and their answers from StackOverflow[10] associated to the tag 'Java' using the scripts described in Chapter 4. This data consisted of 4041 records (500 questions, 3541 answers) from 2091 different StackOverflow users. From this data, we created 4 initial designs patterns listed in Table 3 and assigned them sample discussions.

**Table 3.** Initial design patterns created for the user study.

| Pattern name | Problem | Solution | Related Pattern |
|---|---|---|---|
| Reference Provider | The user is trying to address a particular functionality and it is looking for references to services that can address the problems. | Provide a reference to an external service or utility that can provide the needed functionality. | N/A |
| Trusted Advisor | The answer seeker is looking for advice regarding the recommended or appropriate way to implement a task or | To answer these types of questions, the answerer should start by providing the recommended answer to address the problem or situation. Then, | N/A |

| | complete a set of instructions. The answer-seeker might know the answer or a set of answers but might be looking for confirmation. | the answered should provide some level of reasoning behind the benefits of the selected answer. | |
|---|---|---|---|
| Educator | The answer seeker is looking for an explanation of a topic, error message or code. | poo9 | Comparator |
| Comparator | The answer seeker is trying to determine the difference between two or more topics. | The answerer should provide a comparison between the different options submitted by the answer seeker. | Educator |

### 5.1.3   Study Session Components

Each study session consisted of five components. First, participants completed a consent form and a screening questionnaire. Then we gave them a brief tutorial and asked them to complete twelve (12) tasks for each system. Finally, we asked participants to complete an exit questionnaire followed by a summary interview.

### 5.1.3.1  Consent Form and Screening Questionnaire

Prior to starting the user study, we asked participants to complete a consent form that instructed participants about the overview of the study and the activities they would be completing. It also informed them of any potential risks and how they will be compensated for their participation. Immediately after completing the consent form, we asked participants to complete a screening questionnaire. The screening questionnaire served two purposes. First, it allowed us to collect demographic and background information about the participants. Second, it allowed us to determine, prior to starting the study, whether the participant met the inclusion criteria for participation. The Institutional Review Board (IRB) approved the study with reference number PRO14040539. The consent form was a required step in IRB procedures for performing studies with human subjects. All participants provided informed consent.

### 5.1.3.2  System Tutorials

Prior to commencing the experiment, we performed a randomization assignment and assigned participants to start with one of the two systems: MB or Q&A. Had any participant not met the inclusion criteria, we would have stopped the study and we would not have reused neither their position in the registration nor their random assignment. At completion of the user study, 17 participants started with Q&A followed by MB and 15 participants started with MB followed by Q&A.

Depending on the system assigned to begin with, we gave participants a 10-minute tutorial on the basic usage of the system. The tutorial covered how to use the system to complete a set of basic operations. These basic operations included how to navigate the discussions, how to search for discussions, how to create a question, and how to create an answer. For the case of

MB, we additionally explained what design patterns were to participants and how to use them to discover discussions and to associate them to discussions. One challenge we had was with trying to close the learning curve gap of performing tasks using MB. To address this, in addition to giving participants a demonstration of the features in MB, we also asked them to complete two practice tasks using MB prior to starting the study tasks. For the case of Q&A, we let participants interact with the application following the tutorial.

### 5.1.3.3 User Study Tasks

Following completion of the tutorial, we asked participants to complete a set of tasks (APPENDIX A) that tried to simulate the scenarios of how students would have used the systems in an educational domain. Each task consisted of a brief description and asked participants to use the currently assigned system to complete a particular action or actions. Because some tasks had deterministic answers while others were open ended, participants were asked to either write down the identifier of the discussions that solved the task or write down what they did to complete the task using the system. In addition, using a 5-point Likert-scale, participants were asked to rate the perceived difficulty of completing the task using the current system and how likely the given task is to be performed in an educational setting.

We did not give participants a time limit to complete the tasks but we told them to inform us if any task was not clear or confusing. Once participants completed all tasks using the first system, we switched to the second system where the same steps (i.e., tutorial and tasks) were completed. Participants completed 24 tasks in total for the two systems.

We organized the 12 tasks into four different scenarios: discovering knowledge using keywords, discovering knowledge using timelines, using patterns for knowledge reuse and using

patterns for knowledge discovery.   Participants completed the tasks in the same order as listed in APPENDIX A.

### (a)   Discovering knowledge using keywords

Keyword searching is the most basic functionality that knowledge discovery systems must support.   For these tasks, participants used the systems to find discussions for particular topics and keywords.  We defined three tasks for discovering knowledge.  Task 01 asked participants to find a question about a given keyword. Task 02 asked participants to find the most popular discussion for a given keyword. Task 03 asked participants to first find an unanswered question about a given keyword and then find related discussion suggested by the system.

### (b)  Discovering knowledge using timeline

These tasks asked participants to discover discussions using temporal constraints.  Task 04 asked participants to find any discussion that occurred during a given timeframe.  Task 05 asked participants to find discussions related to a given keyword that occurred during a given timeframe. Task 06 asked participants to find the most recent unanswered question.

### (c)   Using Patterns for knowledge reuse

For these tasks the goal was to find out if given an open ended scenario that required knowledge reuse,  would participants have considered using design patterns to arrive to their solution?. These tasks were opened-ended particularly because the Q&A system did not have the notion of design patterns; however, the tasks could give us some insight into what existing functionality or behavior participants did to solve them.  Similar to before, we defined three tasks to evaluate

using design patterns for knowledge reuse. Task 07 asked participants to find examples of discussions whose answer provided recommendations, gave advise or suggested the use of references. Task 08 asked participants to modify a given discussion such that other students looking for similar approaches to answering the question could have referenced it. Task 09 asked participants to submit a given answer to an unanswered discussion and then they were asked to modify the discussion such that other students could reuse how the answer to the question was presented.

### (d) Using Patterns for knowledge discovery

Similar to the previous scenario, for these tasks participants were given open ended scenarios that required knowledge discovery. The goal was to find out whether participants in MB used design patterns as their solutions or what other behavior participants did. Task 10 asked participants to find discussions based on the approach taken to answer them. Task 11 asked participants to refer to a particular discussion and then find other discussions whose approach to answering it was similar. Task 12 asked participants how they can find recommended ways to answer discussions.

### 5.1.3.4 Exit Questionnaire

We gave participants an exit questionnaire after completing all 24 tasks with both systems. The exit questionnaire assessed their experience with discussion forums and Q&A systems, their experience using MicroBrowser, their perception on how MicroBrowser compares to a traditional Q&A system, their feedback on using MicroBrowser in an educational setting, and general feedback on MicroBrowser.

### 5.1.3.5 Summary interview

The final component in the user study consisted of a brief interview. During this interview, we asked participants about their experiences with the systems, in particular what they liked or disliked about the systems. We also wanted to get their candid opinion about the introduction and benefits of design patterns.

### 5.1.4 Outcomes

For each component of the user study, we evaluated several outcomes. Table 4 summarizes the components and their respective outcomes. Likert-scale were used to scale responses when completing each task and on the questionnaire. We also captured system usage information in the activity logs for both systems by capturing several metrics such as mouse clicks, keyword searches performed, pages visited and type of page element viewed.

**Table 4.** Outcomes from each component of the user study.

| | |
|---|---|
| Informed consent and screening questionnaire | Consent to participate on the user study, age range, gender, university affiliation, programming experience, and education background |
| Tasks | Perceived difficulty to complete task using system, time to complete each task and likelihood of performing the task in an educational domain |
| Exit Questionnaire | Experience with Q&A systems, Likert-scale responses on effectiveness and easiness of features |

| Summary interview | General comments on the systems and the integration of design patterns |
|---|---|

## 5.2    STATISTICAL ANALYSIS

We used descriptive statistics to summarize the data, including mean (*M*) and standard deviation (*SD*) or median (*Mdn*) and interquartile range (*IQR*) for normally or non-normally distributed continuous data, and frequencies and percentages for categorical data.  Evaluation system data were summarized using descriptive statistics including *M*, *SD*, *Mdn*, and *IQR*.   Frequency distributions were used to summarize count data.  Mann Whitney test was used to determine differences in years of experience and years of Java programming experience between males and females.  Wilcoxon Signed-Rank test was used to determine differences in responses for the tasks, measured using a Likert-scale, between MicroBrowser and the traditional Q&A system. Wilcoxon Signed-Rank test is a non-parametric test (that is, the test does not assume normal distribution of the data) that is commonly used to evaluate Likert-scale data.  McNemar's Chi-square test was used to determine associations between systems and success rate.   Pearson's correlation coefficient or Spearman's rank-order correlation was used to measure the strength of the association between two variables of interest. Rank biseral (Sommers'D) correlation was used to compute correlations between binary values and Likert-scale responses. Simple linear regression or multiple linear regression was used to determine the association between an outcome and independent variable(s).  K-means clustering was performed to classify participants

based on their knowledge browsing and system usage behavior. All analyses were performed

with R Software (version 3.1.3) and all p-values were two-sided with a significance level of 0.05.

## 5.3    RESULTS

### 5.3.1   Participants Demographics

We recruited 32 students to participate in the user study. There were 6 (19%) female and 26

(81%) male participants (Figure 30, on left). Among the participants, we had 12 (38%) students

in the 18-21 age range, 13 (41%) students in the 22-25 age range, 6 (19%) students in the 26-30

age range and 1 (3%) student in the 31+ age range (Figure 30, on right).
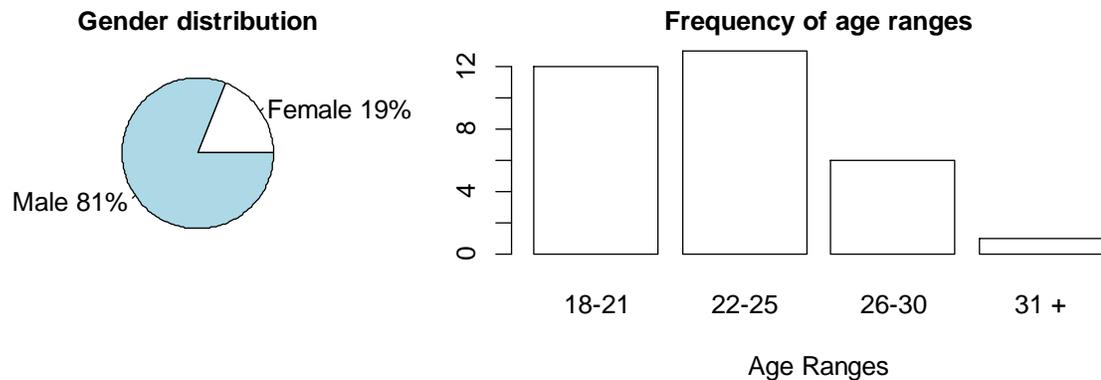
**Figure 30.** Gender distribution and frequency of age ranges for study participants

All 32 participants were students at the University of Pittsburgh. Among the participants,

we had 13 (41%) undergraduate and 19 (59%) graduate students. The average years of college

experience of participants was 4.5 years ($SD = 3.0$, $M_{undergraduate} = 1.8$ years, $SD_{undergraduate} = 1.4$,

$M_{graduate} = 6.3$, $SD_{graduate} = 2.3$). For major of study, 25 students were in Computer Science (78%)

followed by 3 students in Computer Engineering (9%) and one in each of Artificial Intelligence, Information Science, Computer Science & Philosophy and Telecommunications.

Prerequisites to participate in the study included being computer literate and to have some programming experience. For computer experience, all 32 participants considered themselves computer literate. That is, they were familiar with using and interacting with a desktop computer. In addition, participants had an average of 5.1 years of programming language experience ( $SD$ = 3.1) and in particular , an average of 3.2 years of Java programming experience ($SD$ = 2.3). A Wilcoxon rank sum tests indicated no statistically significant difference between males and females in neither the years of programming experience ($Mdn_F$ = 6.0, $IQR_F$ = 4, $Mdn_M$ = 4.2, $IQR_M$ = 3), $Z$ = 68, p = 0.645, nor the years of Java programming experience ($Mdn_F$ = , $IQR_F$ = , $Mdn_M$ = , $IQR_M$= ), $Z$ = 81, p = 0.903.

With respect to previous experience with discussion forums and Q&A sites, all 32 participants were familiar with using discussion forums and Q&A sites such as StackOverflow and Yahoo Answers!. Table 5 presents participants' experience with specific functionalities in Q&A sites.

**Table 5.** Participants' experience with using Q&A sites

| Activity | Has experience | |
| --- | --- | --- |
| | Count | % |
| Browsed questions and answers | 32 | 100 |
| Created questions | 12 | 38 |
| Submitted an answer | 11 | 34 |

112

| | | |
|---|---|---|
| Voted on answers | 10 | 31 |
| Voted on questions | 15 | 47 |
| Viewed their leaderboard | 16 | 50 |
| Created an account | 7 | 22 |

Results show that 16 (50%) participants had used a Q&A system as part of their courses. For those participants that had used a Q&A system as part of their courses, we asked them using a 5-point Likert-scale [1 − Strongly Disagree, 5 − Strongly agree] to what extent they found them useful. Participants agreed that using a Q&A system in a class was helpful for learning course content ($M = 3.6$, $Mdn = 4$). For participants that had not used a Q&A system in any of their classes, using a 5-point Likert-scale [1 − Strongly Disagree, 5 − Strongly agree], we asked them to what extent they would find useful a Q&A system for learning course content. Respondents agreed that they would find useful the use of a Q&A system to learn course content ($M = 4.3$, $Mdn = 5$).

### 5.3.2   GOAL 1: To determine if MicroBrowser promotes peer learning and knowledge discovery and whether students were able to discover and reuse knowledge efficiently.

For this goal, we analyzed perceived task difficulty, likelihood of performing a task, and the time taken by participants to complete the tasks.

**5.3.2.1 Perceived Difficulty and Likelihood of Performing a Task**

Our aim in developing MB was to facilitate peer learning and knowledge discovery and reuse. To investigate this aim, participants were asked to assess the perceived difficulty of performing the tasks. Then, the perceived difficulty of performing the equivalent task with each system was compared. Possible order-effects affecting the results were assessed. Finally, we asked participants to rate the likelihood of performing a similar task in an educational setting to ensure that such tasks were appropriate. We captured the responses using a 5-point Likert-scale (1-Very Easy to 5-Very Difficult). To reduce noise and increase power for the analysis, we summed the Likert-scale item responses for the tasks by their scenario.

Wilcoxon signed rank tests were performed to evaluate the responses to the perceived difficulty of completing the four scenarios. For the scenario of knowledge discovery using keywords, test results indicated that participants felt it was easier to complete the tasks using MB (*Mdn* = 4.0, *IQR* = 1) than using Q&A (*Mdn* = 6.0, *IQR* = 3), Z = 398, p < .001. For the scenario of knowledge discovery using timeline, test results indicated that participants felt it was easier to complete the tasks using MB (*Mdn* = 5.5, *IQR* = 2) than using Q&A (*Mdn* = 7.0, *IQR* = 2), Z = 283, p < .001. For the scenario of using patterns for knowledge reuse, test results indicated that participants felt it was easier to complete the tasks using MB (*Mdn* = 6.0, *IQR* = 3) than using Q&A (*Mdn* = 8.0, *IQR* = 4), Z = 317, p < .001. For the scenario of using patterns for knowledge reuse, test results indicated that participants felt it was easier to complete the tasks using MB (*Mdn* = 7.0, *IQR* = 4) than using Q&A (*Mdn* = 8.0, *IQR* = 4.2), Z = 338, p < .03.

We then looked at order effects to determine if starting the experiments with either MB or Q&A affected the perception of difficulty. Results of the Mann-Whitney tests for all scenarios (Table 6) indicated that when participants started the study using MB, the perception

of difficulty when using Q&A was statistically significant higher than when using MB. These results suggest that order effects are small or not present.

We did not find statistically significant differences in the responses to task difficulty when Q&A was first, however, trends in the data suggest that participants felt that MB was easier to use to complete all scenarios except for the knowledge discovery using patterns scenario.

**Table 6.** Order effect on perception of task difficulty.

| Task scenario | Q&A | | MB | | p-value |
|---|---|---|---|---|---|
| | *Mdn* (IQR) | *M* (SD) | *Mdn* (IQR) | *M* (SD) | |
| **Knowledge discovery using keywords** | | | | | |
| MB, Q&A | 8 (3) | 7.5 (1.8) | 4 (2) | 4.4 (1.2) | < 0.01 |
| Q&A, MB | 6 (2) | 4.9 (1.8) | 4 (1) | 4.9 (1.8) | .28 |
| **Knowledge discovery using timeline** | | | | | |
| MB, Q&A | 7 (2) | 6.8 (2.1) | 5 (2) | 5.2 (1.1) | < 0.01 |
| Q&A, MB | 7 (2) | 7.1 (1.7) | 7 (2) | 6.7 (1.8) | .42 |
| **Knowledge reuse using patterns** | | | | | |
| MB, Q&A | 8 (3) | 7.8 (2.2) | 6 (1) | 5.9 (2.1) | .01 |
| Q&A, MB | 9 (5) | 9 (3.1) | 8 (5) | 7.9 (2.9) | .26 |
| **Knowledge discovery using patterns** | | | | | |
| MB, Q&A | 9 (3) | 8.5 (2.3) | 6 (3) | 6.2 (2.3) | .01 |
| Q&A, MB | 8 (6) | 7.5 (3.6) | 9 (3) | 7.7 (2.3) | .97 |

Results for the likelihood of performing the tasks in an educational setting show that participants felt that all tasks were likely ($Mdn_{\text{Task 01-12 except 08}} = 4$, $Mdn_{\text{Task 08 MB}} = 3$, $Mdn_{\text{Task 08 Q\&A}} = 4$) to be perfomed in an educational setting.

### 5.3.2.2 Discussion

Based on the results, participants found MB to be easier to complete all of the scenarios when compared to Q&A. Results show that participants found the tasks of knowledge discovery using keywords to be the easier tasks overall when using MB. For the scenario of knowledge discovery using timeline, participants benefited from using the knowledge timeline to narrow discussions as well as the ability to filter by both keyword and timeline simultaneously. When using Q&A, participants could only sort by recent discussions but had to page thru the discussions until reaching the desired date.

For the scenario of knowledge reuse using patterns, we wanted to see if participants could use the systems to identify instances of reusable knowledge. Using MB, participants were able to use the pattern browser and from there refer to the sample discussions. In the case of Q&A, participants used different methods, such as performing keyword searches and then reading the details of identified discussions. Sometimes participants simply opted for selecting the discussion that matched a keyword search. Participants benefited by the use of design patterns. Reviewing the data for Tasks 08-12, we noted that participants made use of patterns while completing the tasks in 66%, 50%, 81%, 78% and 77% of the times, respectively. For the case of Q&A, participants tended to associate a tag to create a classification of the discussions.

**5.3.2.3 Time Completing Tasks Analysis**

While task difficulty gave us a qualitative measure of the effectiveness of MB, we looked at the time participants spent while completing the assigned tasks as a quantitative measure. We expected time to present us a challenge because of the learning curve with MB that could bias the results towards Q&A. However, results from the time analysis can help us identify potential areas of improvements.

The tasks for knowledge discovery using keywords and timelines scenarios can provide us a way to compare usage behavior between the two systems. On the other hand, the tasks for knowledge reuse and knowledge discovery using patterns scenarios were open-ended tasks. Therefore, time measurements captured the time participants spent trying to discover how the system can facilitate the completion of the tasks. As an example, participants instead had to depend on the available functionality to complete the tasks since the concept of design patterns was not available in Q&A.

Wilcoxon-signed rank tests were performed to evaluate the time to complete the tasks of the four scenarios using MB and Q&A systems. For the scenario of knowledge discovery using keywords, test results indicated that participants completed the tasks faster using MB (*Mdn =* 40.2*, IQR =* 20, *M =* 42.2, *SD =* 21.5) than using Q&A (*Mdn =* 85.5, *IQR =* 64, *M =* 82.5, *SD =* 43.1), Z = 474, p < .001. For the scenario of knowledge discovery using timeline, test results indicated that participants completed the tasks faster using Q&A (*Mdn =* 47.7, *IQR =* 27, *M =* 48.3, *SD =* 21.8) than using MB (*Mdn =* 67.7, *IQR =* 30, *M =* 71.5, *SD =* 30.3), Z = 80, p < .001.

We did not find statistically significant differences for the time participants took to complete the tasks for the scenarios of knowledge reuse using patterns (Z = 200, p = .23) or knowledge discovery using patterns (*Z =* 326, p = .24). Looking at the trends in the data, we

noted that participants appeared to spend less time completing tasks for the knowledge reuse using patterns with Q&A (*Mdn* = 81.5, *IQR* = 40, *M* = 86.2, *SD* = 34.9) than MB (*Mdn* = 81.7, *IQR* = 91, *M* = 113.4, *SD* = 85.3).  On the other hand, for the scenario of knowledge discovery using patterns, the data shows a trend to suggest that participants spent less time to complete the tasks using MB (*Mdn* = 81.5, *IQR* = 40, *M* = 86.2, *SD* = 34.9) than Q&A (*Mdn* = 81.5, *IQR* = 40, *M* = 86.2, *SD* = 34.9).

**5.3.2.4 Discussion**

Interestingly, for the scenario of knowledge discovery using keywords, we found participants took less time to complete the tasks when using MB.  Since this scenario included common tasks for Q&A systems in general, the fact that data shows MB quicker to complete the tasks than Q&A suggests validation of our approach.  If we consider the learning curve effect affecting task time, there are real possibilities of further improvement. We believe that features in MB such as displaying connections to similar discussions as well as adding a node labeled *Related Discussions* in the Discussions details view facilitate the process of finding related discussions. The Q&A system only displayed the list of related discussions on the discussion details but it was not easy to notice.

For the scenario of knowledge discovery using timeline, we expected the knowledge timeline functionality in MB to provide an advantage.  However, our results showed that participants completed the tasks faster using Q&A.  Features that could explain this include the fact that Q&A sorted discussions by recent updates by default resulting in less steps required to find recent discussions.   When filtering discussions based on keywords and timeline, participants were able to perform the search and then scroll thru the pages until the

corresponding timeframe came into view when using Q&A. After filtering discussions by a keyword, the number of pages required to display the results was small. Thus, there were not a lot of result pages participants had to browse to find discussion when using Q&A. However, with a larger set of discussions there might be more browsing required. In contrast, when using MB, participants had to first change the timeline and then perform the keyword search. We believe that the learning curve effect might have affected the performance of participants when using MB since participants had to combine two search methods to filter discussions. For Q&A, participants had the ability to simply click on the *Unanswered questions tab* which was already sorted by recent activity. Using MB, participants had to use the *Knowledge Timeline* to select different time ranges until they could find an unanswered discussion. This activity required several attempts to modify the time ranges affecting the time to complete the task. Therefore, an example of future improvement for MB is to allow users the ability to filter discussions by their current state or by presenting more sorting capabilities to facilitate knowledge exploration.

For the scenarios of knowledge reuse and knowledge discovery using patterns, we did not find statistically significant differences between Q&A and MB. These results suggest that participants were able to find alternative methods using each system. For example, with MB, participants were able to use design patterns as reusable components while for Q&A, students performed approaches such as performing keyword searches or using the *Hot Questions* functionality. Since the tasks were of different levels of difficulty, it was not feasible to analyze learning curve improvement between tasks.

### 5.3.3 GOAL 2: To determine if MicroBrowser facilitates completion of knowledge discovery and reuse scenarios and if students understood the concept of design patterns.

For this goal, we focused on success rate for knowledge discovery and on integration of design patterns.

### 5.3.3.1 Success Rate for knowledge discovery using Keywords and Timeline

Success rate can help us understand whether our solution facilitates completion of common tasks performed when working with Q&A systems. In particular, we were interested on determining whether our new solution did not affect the ability of participants to complete tasks using the MB system. For the knowledge discovery using keywords and timeline scenarios' tasks, we measured success rate as the percentage of tasks that participants reported a valid answer. For the knowledge discovery and reuse scenarios' tasks, we focused on MB and defined success rate as whether participants integrated the use of design patterns in finding a solution.

McNemar's tests were performed to compare the success rate of completing the tasks for the scenarios of keyword and timeline search scenarios using the MB and Q&A systems. Results of the test indicated a significant difference in success rate suggesting that for the scenario of knowledge discovery using keywords, the success rate was higher with MB ($success_{MB} = 82\%$) than Q&A ($success_{Q\&A} = 68\%$), $\chi^2(1) = 3.8$, $p = .05$. Results of the tests also indicated that for the scenario of knowledge discovery using timeline, no significant difference in success rate was noted, $\chi^2(1) = 0.0$, $p = .94$.

Given the availability of the timeline search view in MB, we wanted to understand what could have affected the success rate for the timeline search scenario (Tasks 04 – 06). Task 04

asked participants to filter the discussions that had an activity during a specific range of days. Reviewing the video recordings and data logs for participants that provided an incorrect response, we found that participants followed the correct steps except that while the correct month and day were selected for the starting date, the year was not correct. Since the year attribute of the date range was incorrect, the system returned discussions that were outside the correct range. Future work for the MB system should focus on improvements on the process of selecting the date ranges. Currently, we present a date picker dialog that put more emphasis on the month and day selection but does not require users to change the year. Instead, we could redesign the date selection to first ask the students to either confirm or select a starting year and then present them the month and date selections. Another possible alternative could be giving students a step-by-step date selection option that asks for each date componet specifically.

## 5.3.3.2 Association between success rate and perceived task difficulty

We assessed the association between success rates and how easy participants thought the MicroBrowser systems was. To perform this analysis, we computed a rank-biserial correlation between the success rate (binary, 1 – success, 0 - failure) and the 5-point Likert-scale responses to task difficulty (1 – Very easy to 5 – Very difficult) for Tasks 01-06. Rank-biserial correlation results show a very strong negative correlation for Task 01 ($r(30) = -0.74$), a weak negatve correlation for Task 02 ($r(30) = -0.27$), a strong positive correlation for Task 03 ($r(30) = 0.50$), and no or negligible correlation for Task 04 ($r(30) = -0.12$), Task 05 ($r(30) = 0.02$) and Task 06 ($r(30) = -0.07$). A negative correlation means that participants that successfully completed the task were likely to perceive the task as easy to perform using the current system in use.

**5.3.3.3 Success Rate for the Knowledge Reuse and Discovery Patterns.**

For the scenarios of using patterns for knowledge reuse and knowledge discovery (Tasks 09-12), we focused on MB and defined success rate as whether the participants integrated the use of design patterns. This integration could be either identifiyng an appropriate pattern or using a pattern to perform knowledge discovery. Results show when identifying the solution for Task 09-12, the percentage of participants that used patterns was 50%, 81%, 78% and 77% respectively. These results demonstrate the even though it is a new concept, over 50% of the participants integrated design patterns as part of the solution.

**5.3.3.4 Participants created design patterns**

A goal of the integration of design patterns is to facilitate knowledge creation. To that end, we asked participants if they could suggest additional design patterns that we could add to MB. Even though this was an optional question, we received 11 suggestions for proposed design patterns (Table 7). Reviewing the list of design pattern suggestions, we see that 8 of the suggested design patterns have a relationship to an educational domain. In particular, 2 design patterns were about questions on homework or class material clarifications (Assignment/Homework, Tutorials), 3 design patterns were for programming problems (Alternatives, Debug, Summary), 3 design patterns were about repeating questions and also for finding good and bad discussions (Previously answered, Well answered/Well asked, Bad questions/subcategories). Finally, one design pattern was for using graphs to answer questions (Graph), one for identifying cool stuff (Cool Stuff) and one to capture social aspects (Power users).

**Table 7.** Participant suggestion list of design patterns to add to the MB system.

| Pattern name | Pattern description | Pattern solution |
|---|---|---|
| Power user | User with many correct answers and has answered | Helps the user to identify/find users with reputations |
| Alternatives | Asking for alternative to a well-known program/service | Provide alternatives to program/service along with comparison to the well-known option |
| Assignment / Homework | Questions regarding specific projects/ or in class discussion topics or other classroom related stuff | Help students find questions to specific dates/times of tests or projects and other course-specific info |
| Cool Stuff | A discussion on something cool, just for fun | Not available |
| Tutorials | Discussions that pose challenges or questions | Reference to outside code or just directly insert code |
| Debug | Fixing or describing methods for a particular section of code | Give pseudo code fragment with a related piece of code coupled with an explanation. |
| Previously answered | Similar to reference, however exactly same question with answers | You should share a link with the user and give some details, so he/she can be interested in access the link |
| Well answered / Well Asked | Questions/answers that are generally agreed to be relevant and presented in a good manner | Not available |
| Bad questions subcategories | The asker is asking poorly because … | This question can be better by doing … |
| Summary | Summarize some info, like the possible errors in programming | Browse the previous answers or correct their knowledge to form a summarized answer. |
| Graph | Those questions that contain graphs or diagrams | Answers should include graphs, figures or diagrams |

### 5.3.3.5 Integration of Design Patterns

The majority of participants found the use of design patterns effective (81% agree) and most importantly easy to learn (78% agree) (Figure 31). This was an important finding since the aim of the user study was to determine if participants could understand the value of design patterns. Participants were ecstatic about their capability and usefulness as noted by the following comments: *"Pattern browsing is a great concept", "I really loved the pattern browser. Just a brilliant idea"* and *"Pattern browser, it is very useful to see "trusted advisor". If I want to see reliable answers and "reference" and "comparator" are very useful."*

123
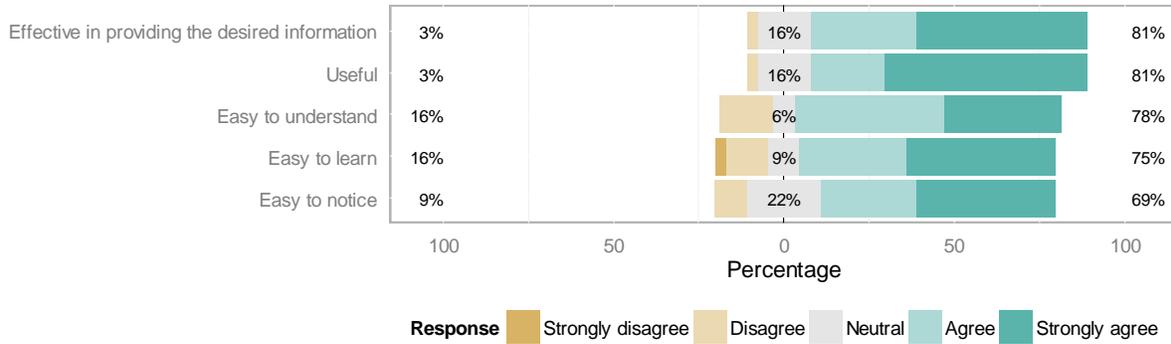
**Figure 31.** Participants' opinion about design patterns (5-point Likert-scale)

### 5.3.3.6 Discussion

Results from the success rate analysis show that students were able to complete the majority of the assigned tasks. When evaluating the success rate of the knowledge discovery using keywords scenarios, we found that participants had a higher success rate when using the MB system than the Q&A system. Associations between success rate and perceived task difficulty were also evaluated. The results suggest that participants were more likely to rate Tasks 01 and 02 as easier if they answered the question correctly. For the using patterns for knowledge discovery and reuse scenarios, we found that more than half of the participants made use of patterns while completing the scenarios' tasks.

Participants liked the idea of design patterns and found them effective at providing the desired information and useful. We were extremely encouraged by the good acceptance of the concept of design patterns by the participants.

**5.3.4   GOAL 3: To determine if MicroBrowser improved user experience for browsing Q&A discussion threads with its novel design features of color coding, node sizes, and keyword and timeline search, and discussion overview.**

For this goal, the focus was on the responses to the questionnaire about MB features. Results presented in Table 8 show that participants found the MB features to be easy to understand, useful and effective. Participants agreed with the statements that the use of color-coding to demonstrate popularity (*Mdn* = 4) and the use of different node sizes to represent answer counts (*Mdn* = 4) were easy to understand.

Participants also strongly agreed with the statement that the keyword search feature when using MB was useful and effective at providing the desired information (*Mdn* = 5). Since keyword search is one of the most common activities performed, these results are encouraging. Participants agreed with the statement that the timeline search in MB was useful and effective to providing the desired information (*Mdn* = 4). From the results of the success rate analysis in section 5.3.3.1, we believe that there is room for improvement in the design of the timeline search in the MB system. Participants also agreed with the statement that the discussion details view was useful (*Mdn* = 4) and effective (*Mdn* = 4) at providing the desired information.

In addition to the predefined questions in the questionnaire, we also asked participants about the features they liked and disliked the most about the MB system. From the responses, the features that participants liked the most were the integration of design patterns and the pattern browser (56%), the visual representation of discussions (47%), visualizing relationships (13%), the use of color to encode popularity (25%), the use of shape encoding (25%), and the knowledge timeline (22%). Among the features participants disliked the most included the

initial learning curve (13%), the knowledge timeline (13%), the slow response of the system (9%) and the use of patterns (9%).

Participants also identified some usability issues and made recommendations for improvements to MB. Among the recommendation were to add sorting capabilities, to improve the knowledge timeline by using a slider, to add filtering by users or question state, to associate more patterns to questions, to add actions on right-click menus and to design the system to allow participation for people with visual disabilities.

**Table 8.** Participants' opinion about MicroBrowser features and functionality. 5-point Likert-scale [1 – Strongly Disagree, 5 – Strongly Agree]

| Question | Median | IQR |
|---|---|---|
| It was easy to find answered and unanswered questions (node shape) | 5 | 1.0 |
| The use of different colors to demonstrate popularity was easy to understand | 4 | 1.0 |
| Using different sizes to show answer counts was easy to understand | 4 | 1.0 |
| Keyword search was useful | 5 | 1.0 |
| Keyword search was effective at providing desired information | 5 | 1.0 |
| Timeline search was useful | 4 | 1.0 |
| Timeline search was effective at providing desired information | 4 | 1.0 |
| Discussion details view was useful | 4 | 1.0 |
| Discussion details view was effective at providing desired information | 4 | 1.5 |

**5.3.4.1 Discussion**

Results show that participants agreed that finding answered and unanswered questions was easy using the MB system. In addition, MB facilitated the identification of discussions based on their answered state by their shape representation. Twenty-five percent of participants cited shape encoding as one of their most liked features. Participants agreed that the use of color-coding to denote popularity and node sizes to demonstrate answer count were easy to understand. Participants also felt that performing keyword and timeline searches were both useful and effective at providing the desired information. Finally, participants also indicated that the discussion details view in the MB system was both useful and effective. In addition to these features, participants also reported to like the MB system because of the integration of design patterns, the pattern browser and the use of visual representations of the discussions. However, some participants also indicated to dislike the learning curve when using the MB system.

**5.3.5 GOAL 4: To determine if MicroBrowser was easy to use and whether students would consider integrating it into their educational setting.**

For this goal, we analyzed the questionnaire responses of participants to the questions of how completing common activities between MB and a traditional Q&A system compared. We then assessed the perceived benefit to participants on using MB in an educational setting. Finally, we analyzed their responses to the overall ease of use of MB.

**5.3.5.1 Comparing MB to a Traditional Q&A System**

We analyzed the responses after asking participants to indicate how MB features compare to those of a traditional Q&A system. In particular, we were interested on learning for what type of activities either MB or a traditional Q&A system was better. As noted in Figure 32 over 90% of participants felt that MB was better at finding information based on their temporal properties. In addition, approximately half of the participants felt that MB was better because it was easier to use. Also, approximately half of the participants felt that finding questions based on their state of either answered or unanswered was easier when the system in use was MB. Several factors might have affected these findings. After reviewing the summary interviews and survey feedback, we noted that several participants indicated missing the ability to sort discussions in MB by different criteria while a participant made a comment that the shape of the nodes was too similar (diamond vs. hexagon) to each other.
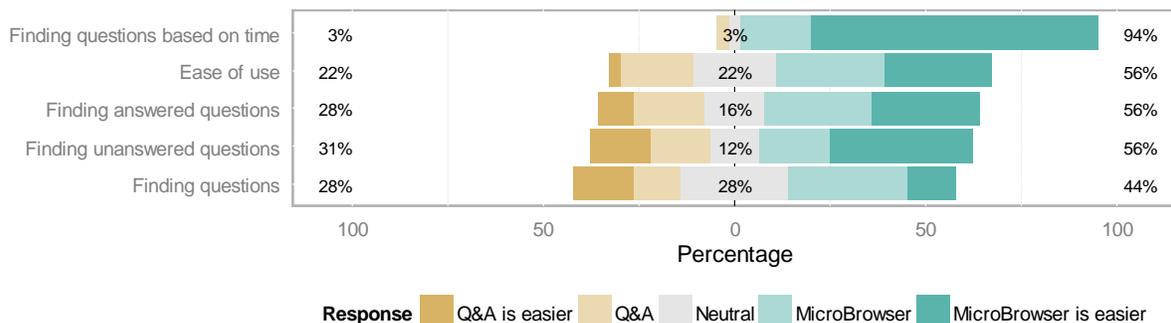


**Figure 32.** Descriptive statistics summarizing how completing different activities compare between traditional Q&A system and MicroBrowser.

128

**5.3.5.2 Using MicroBrowser in an Educational Setting**

We assessed questionnaire responses to determine whether participants have interest on using MB as part of their education and what benefits they believe to obtain from its use. As noted in Table 9, participants agree with being able to use MB as part of their courses. Participants also agreed with the statement that using MB could help them with knowledge sharing between classmates and with knowledge generation. Participants also agreed that MB could motivate them to ask and answer more questions, which can lead to an increased participation rate.

**Table 9.** Descriptive statistics for Using MicroBrowser in an educational setting. 5-point Likert-scale (1 – Strongly disagree, 5. Strongly agree)

| Question | Mdn | IQR |
|---|---|---|
| I would like to be able to use MicroBrowser in one of my courses | 4 | 1.3 |
| Using MicroBrowser in a course can help with knowledge sharing between classmates | 4 | 1.0 |
| Using MicroBrowser in a course can help with knowledge creation, for example, by creating new patterns, questions or answers | 4 | 1.0 |
| Using MicroBrowser in a course would motivate me to ask more questions | 4 | 1.5 |
| Using MicroBrowser in a course would motivate me to answer questions | 4 | 2.0 |

### 5.3.5.3 Ease of use of MicroBrowser

Finally, we analyzed the responses to the questions of overall ease of use of MB. On a 10-point Likert-scale, participants felt that MB was easy to use ($M = 7.8$, $SD = 1.2$, $Mdn = 8$, $IQR = 2$).

### 5.3.5.4 Discussion

We started by comparing common activities between MB and a traditional Q&A system. Results showed that the majority of participants felt that the strength of MB was in discovering discussions based on their temporal attributes. Results from this study also show that participants not only agreed with the idea of using MB in one of their courses but they also agree with MB helping them to share knowledge with peers and to create new knowledge. Finally, results show that participants felt that MB was easy to use.

### 5.3.6   Usage Analsys

Automated classification techniques open new possibilities for interaction design such as adapting the system to the user by changing mouse click behaviors or adapting the amount of information displayed for each discussion based on the users' classification. We explored classifications based on usage behavior and demographic characteristics of participants.

### 5.3.6.1 Classifying participants based on usage behavior

One interest of the user study was to find classifications of participants based on how they interacted with MB. We assessed two different behaviors for classification: how participants accessed the details of a discussion and the usage of either a graph or a text-based browsing

scheme to browse discussions. Results from the classification analysis will provide design recommendations that will be useful to improve the MB system in future revisions and for system designers for extending and improving on MB.

**(a) Classifications based on how participants accessed the details of a discussion**

For this classification, our goal was to understand which method participants used to open the details of a discussion. We looked at the activity of opening details of discussions because, rather than merely browsing on the interface, participants explicitly need to click on an item to complete the action. Figure 33 reveals that there are differences among participants on how frequently they accessed the details of a discussion using either the text-based discussion listing or the knowledge graph views.
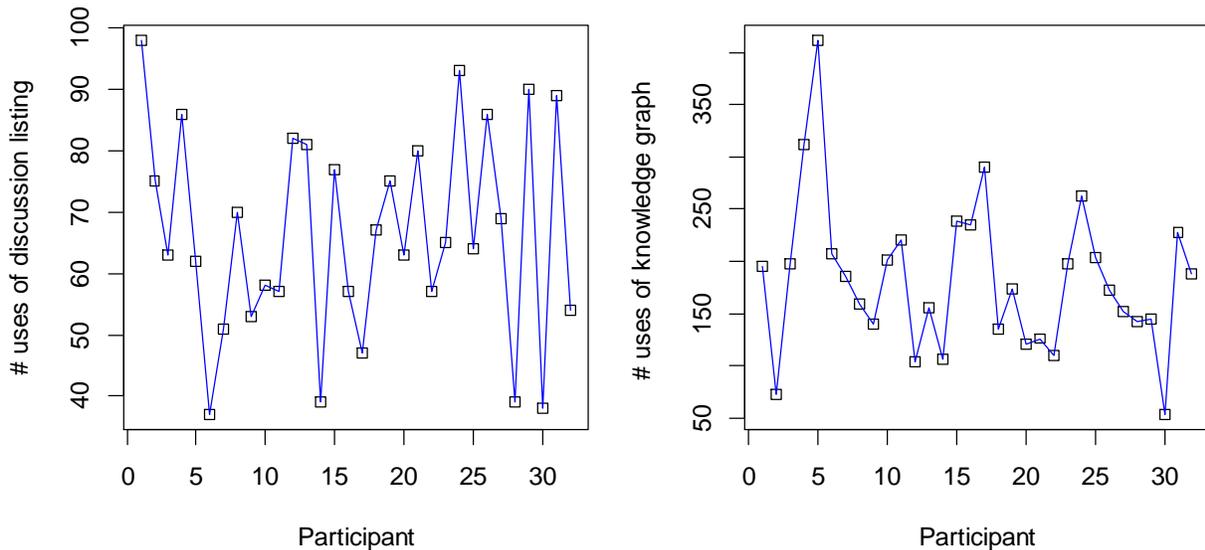


**Figure 33.** Frequency of method used to access discussion details by participant. Discussion listing on left and knowledge graph on right

131

We used k-means clustering, an unsupervised learning technique, to classify participants into groups using three factors based on the number of times participants accessed the details of a discussion either from the overview knowledge graph, from the details knowledge graph or from the discussion listing view. To determine the appropriate number of clusters, we generated a scree plot of the within groups sum of squares by number of clusters as shown in Figure 34. Based on the results from the scree plot, we selected four clusters ($k = 4$) and applied the k-means technique to partition our participants into clusters as shown in Figure 35.
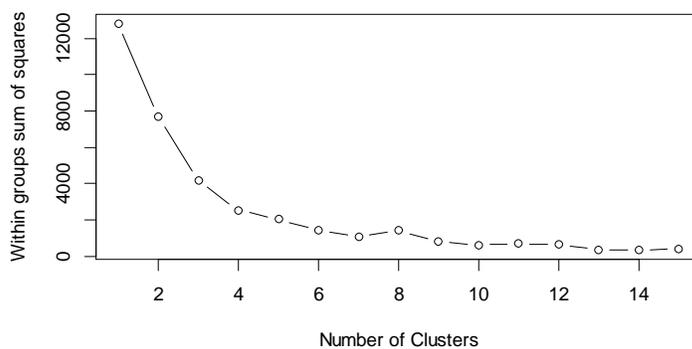


**Figure 34.** Scree plot to identify the number of clusters for classification based on usage behavior

132
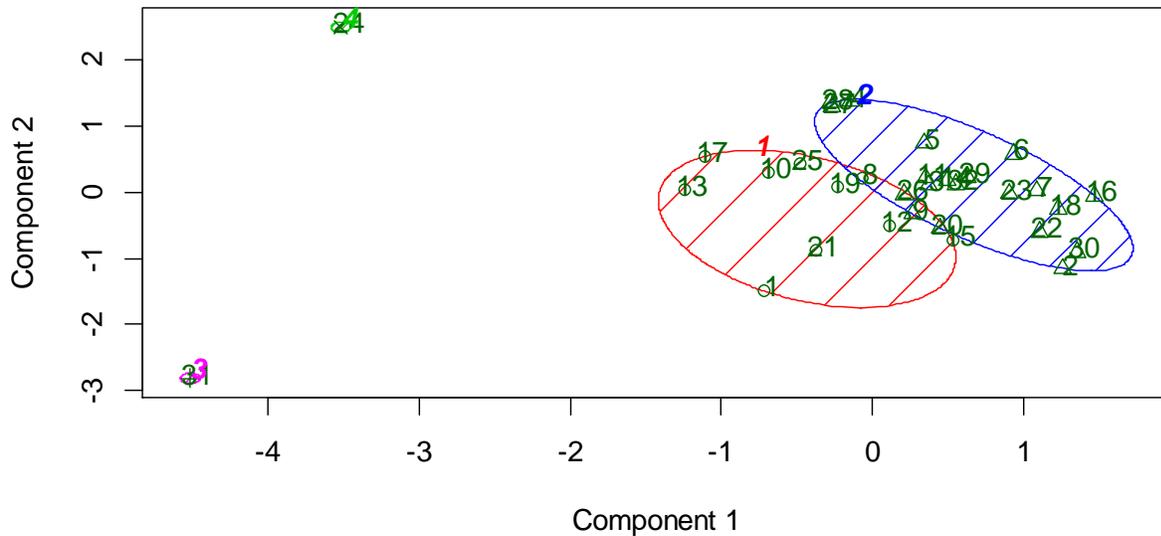
**Figure 35.** Plot of the clusters of participants based on how they accessed discussion details. Numbers on the plot represent participant number.

To better describe the four groups identified, we looked at the details of the means of the members of each group (Table 10). Cluster 1 contained the second largest group of members. The members of this cluster show an increased preference for accessing details of the discussions from the discussion listing view but a higher preference for accessing the details of discussions from the discussion detail graph. Cluster 2 had the largest number of members. Results show that the members of this cluster do not show noticeable differences on the methods used to access the details of the discussion details. Cluster 3 had a single member and the results show this member had a higher preference for accessing discussion details from the discussion listing view. Finally, Cluster 4 had also a single member and this member showed a marked preference for accessing details from the knowledge graph from either the overview or discussion details view. This member also had a very low usage of the discussion listing view.

133

**Table 10.** Cluster means for access discussion details factors

| Cluster | n | Cluster Mean | | |
|:---:|:---:|:---:|:---:|:---:|
| | | **Access item from overview diagram** | **Access item from discussion listing** | **Access item from discussion details graph** |
| 1 | 10 | 6.8 | 14.8 | 26.0 |
| 2 | 20 | 7.2 | 8.3 | 8.0 |
| 3 | 1 | 17.0 | 78.0 | 41.0 |
| 4 | 1 | 22.0 | 2.0 | 64.0 |

From a design perspective, designers can take advantage of this classification to adapt the system to facilitate the access to discussions depending on the user classification. For example, one such modification could be to reduce the number of clicks to access the details of a discussion based on the preferred view. Results show that Cluster 2 that had the most participants did not prefer how they accessed the details of the discussions. However, Cluster 1, the second largest group of participants showed a preference for discovering discussion from the discussion details view. In a sense, these participants tended to follow links to related discussions to discover information. Design implications of these results suggest improvements to the MicroBrowser system to support the different usage behaviors. For example, to reduce the need to jump to other discussions, we can introduce the concept of a single click to display a quick view of the related discussion and a double click to open it. We could also redesign the discussion details to make give more emphasis to the related discussion information such as allocating more space or bigger shapes to them. Another design change we can introduce is the

option to show and hide either the discussion listing view or the knowledge graph to better suite the preference of the participants.

**(b) Classification based on how participants browsed discussions**

The mouse is the preferred input modality for browsing content using the MB system. As participants interacted with the mouse, the MB system generated mouse events of the area or elements that a user is currently focusing on. Using this information allowed us to understand the browsing behavior as participants interacted with the system. For this classification, our goal was to investigate how participants interacted with the MB system using their mouse while exploring the knowledge base content. We wanted to identity three classes of users: Knowledge Graph users, Discussion Listing users and No Preference. For this classification we quantified the number of mouse enter events generated every time a participant placed the mouse either over the details of a node in the knowledge graph or when they placed the mouse over the details of an entry in the discussion listing view. Using these factors, we then performed k-means clustering techniques to classify our participants into the three desired groups (k = 3). As shown in Figure 36, cluster participants had very little overlap to each of the generated clusters.
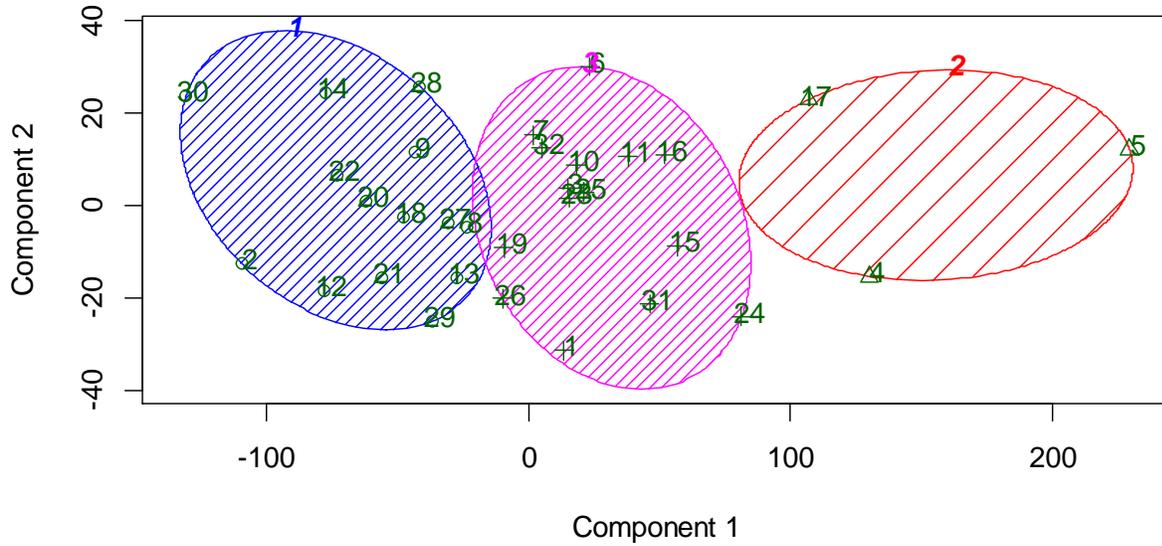
**Figure 36.** Plots of the cluster of participants based on mouse enter events. Numbers on the plot represent participants number.

Table 11 presents the cluster mean results of the k-means clustering algorithm. As noted from the table, there was not a lot of variability among participants concerning the mouse enter events occurring on the discussion listing view. This suggests that participants were equally likely to use the view to browse discussions. The number of mouse enter events generated from knowledge graph had a bigger effect on cluster partitions. Data shows that participants in all clusters had more than double the mean number of events originating from browsing discussions using the knowledge graph than the discussion listing view.

**Table 11.** Cluster means for mouse enter event factors

| Cluster | n | Cluster Mean | |
|---|---|---|---|
| | | List mouse enter events | Graph mouse enter events |
| 1 | 14 | 64.5 | 122.9 |
| 2 | 3 | 65.0 | 338.0 |
| 3 | 15 | 68.3 | 207.1 |

Cluster 1 had 14 members that preferred to use the knowledge graph more often than the discussion listing view, at a lesser degree than participants on the other two clusters do. There were 3 participants on Cluster 2. These participants showed the largest preference for browsing discussions using the knowledge graph. Cluster 3 had 15 members (approximately 46% of the participants) and these participants had the largest mean usage for the discussion listing view than the members of the other two cluster. The preference to use the knowledge graph for Cluster 3 members was much less than Cluster 2 members but more than Cluster 1 members.

Results from this classification show that participants spent the majority of the time discovering information using the knowledge graph. To support this behavior, we would focus design changes on improving access to discussion using their graphical representation. In the previous section, we proposed for example, adding a quick view to allow participants to view information about a discussion without accessing its details. For example, displaying the question text and the accepted answer. We can also find way to reduce mouse travel distance by optimizing the placement of discussions on the knowledge graph either statically or dynamically based on the current discussion on focus (e.g. bringing related discussions closer).

**5.3.6.2 Effects of demographic and experience in system usage**

Our MB system supported two browsing schemes, either graph based navigation or text based navigation. Text based navigation takes place as participants interact with the discussion listing view, while graph based navigation occurs when the participants interact with the knowledge graph view.

To determine the most common navigation area, we assessed mouse events captured as participants browsed discussion node elements on the system. We focused on the two main views of the application: the knowledge graph view and the discussion listing view.

**(a) Are demographic or experience characteristics of participants associated with knowledge graph or discussion listing view usage?**

Data shows that on average, participants browsed discussions using the knowledge graph 72% of the time while 28% of the time participants were browsing discussions using the discussions listing view. As reported in Section 5.3.1, participants in the user study had different demographic and experience characteristics. Therefore, we wanted to investigate whether participants' demographic and experience characteristics were associated with their usage of the knowledge graph or the discussion listing view in a univariate form. We used Spearman rank-order correlation to analyze such association. We did not include age as a characteristic because we found a strong correlation with graduate status ($r(30) = 0.78$, $p < .001$) and college experience ($r(30) = 0.87$, $p < .001$). Results are shown in Table 12.

**Table 12.** Results of correlation analysis between demographic and expertise characteristics of participants and usage of either knowledge graph or discussion listing views. r=Spearman's rank-order correlation, N = 30, gray background p < 0.05.

| Participant characteristics | Using knowledge graph | | Using discussion listing | |
| --- | --- | --- | --- | --- |
| | r | p-value | r | p-value |
| Gender | 0.12 | .51 | 0.01 | .96 |
| Graduate | -0.37 | .03 | 0.11 | .52 |
| College experience | -0.14 | .43 | 0.35 | .05 |
| Programming experience | -0.23 | .19 | 0.07 | .70 |
| B - Created questions | -0.27 | .12 | -0.04 | .79 |
| C - Submitted an answer | -0.08 | .66 | 0.20 | .28 |
| D - Voted on answers | 0.08 | .66 | 0.31 | .08 |
| E - Voted on questions | 0.15 | .40 | 0.25 | .16 |
| F - Viewed their leaderboard | -0.13 | .45 | -0.10 | .57 |
| G - Created an account | -0.28 | .12 | 0.09 | .62 |

Correlation results indicated a moderate negative correlations between knowledge graph usage and graduate status, r(30) = -0.37, p = .03. From these results, we conclude that undergraduate students were more likely to interact more with the knowledge graph. Results also show statistically significant moderate positive correlation between college experience and usage of the discussion listing view, r(30) = 0.35, p = .05. From these results, we conclude that participants with more college experience are likely to interact more with the discussion listing view.

### (b) Modeling usage of knowledge graph and discussion listing views from participants' expertise

The goal was to build models to describe the usage of the discussion browsing method using participants expertise. To define expertise, we considered years of experience and prior experience with Q&A systems. For years of experience, we considered college experience, java experience and programming experience. We looked at the correlation among them and found them to have strong correlation among themselves, $r(30) > 0.66$. Therefore, we opted to use college experience as our experience variable. For prior experience with Q&A systems, we considered all variables from Table 5.

To determine the most important variables to include in our multivariable linear regression models, we performed a stepwise regression model selection. We began by generating a model to describe usage of the knowledge graph view. Results of our stepwise regression analysis (Table 13) suggested a model comprising of the variables create questions, vote on questions and create an account, Adjusted $R^2 = 0.1504$, p = .056. Therefore, 15% of the variability in the usage of the knowledge graph can be explained by these three variables.

**Table 13.** Results from multiple linear regression for usage of knowledge graph.

| Feature | Coefficient | p-value |
|---|---|---|
| (Intercept) | 213.58 | < 0.01 |
| Has experience creating questions. | -48.49 | .14 |
| Has experience voting on questions. | 70.43 | .02 |
| Has experience creating accounts. | -48.80 | .16 |

For the discussion listing view usage, results of the stepwise regression analysis (Table 14) suggested a model that included college experience, voting on answers and creating an account, Adjusted $R^2$ = 0.117, p = .09. These results suggest that the suggested model can explain about 12% of the variability in the usage of the discussion listing view.

**Table 14.** Results of multiple linear regressions for usage of discussion listing view.

| Feature | Coefficient | p-value |
|---|---|---|
| (Intercept) | 213.62 | < 0.01 |
| Years of college experience | -2.24 | .60 |
| Has experience voting on answers. | 69.69 | .04 |
| Has experience creating accounts. | -88.21 | .03 |

### 5.3.7   Pattern Usage Analysis

We analyzed the logs from each participant to obtain descriptive and usage statistics about the use of design patterns by participants (Table 15). The Pattern Selection dialog allows participants to associate patterns to discussions. By doing this, they provide examples of reusable knowledge by the associations between design patterns and discussions. Participants on average opened the pattern selection dialog 2.3 times ($SD$ = 1.8). To view the details of the design patterns participants opened the patterns details. The details about design patterns were useful to obtain information about reusable knowledge and to find examples on how this knowledge has been reused. From the log results, participants opened the details of a design pattern on average 8.1 times ($SD$ = 7.4).

**Table 15.** Pattern usage results

| Activity | *M* | *SD* |
|---|---|---|
| Open pattern selection | 2.3 | 1.8 |
| View pattern details | 8.1 | 3.8 |
| Pattern create – details view | 11.0 | 7.4 |
| Pattern create – main view | 0.0 | 0.2 |

In our study design, we looked at how students could benefit from the integration of design patterns into an interactive knowledge graph system. Results from Section 5.3.3 and Table 15 suggest that participants interacted with design patterns in the process of performing knowledge discovery and knowledge reuse. Understanding this usage can have implication in the design of future revisions on the application. From the results, we noted that while participants had the option of creating design patterns from the overview screen, they were more likely to create the design patterns within the context of the discussions. These results also indicate that participants opened the details of design patterns in addition to having the information in the pattern browser and the pattern selection dialog. In MB, design patterns not only contain general information about patterns, but they also contain references to examples of how the knowledge has been reused. From the activity log, we determine that participants took advantage of this approach to find such examples. We looked at the responses provided by students regarding the selected patterns. From the results, the two most common patterns used were Reference Provider and Comparator. On average, participants selected 2.1 ($SD = 1.0$)

distinct design pattern while completing the scenarios of knowledge discovery and reuse using design patterns.

While participants reported that the concept of design patterns is easy to understand and based on results from Section 5.3.3.3, participants reused design patterns in the process of completing the scenarios' tasks, we believe further work can be done to support complex scenarios that require the use of two or more design patterns. In work by Ginat, (Ginat, 2009) the author recommended scaffolding the process of design pattern identification to support students in the process to develop solutions that combine design patterns.

## 5.4    CONCLUSION

In this chapter, we presented a user study of MicroBrowser, a system that facilitates peer learning and knowledge discovery in classroom settings. Results from our 32-subject user study show reduced difficulty for MB at completing tasks when compared to the traditional Q&A system. More importantly, participants found benefit in the use of design patterns and found the system and its features effective and easy to use.

For this user study, we focused on four main goals to explore the MB system. We also wanted to analyze usage behavior to understand how participants interacted with MB. The first goal was to evaluate if MicroBrowser promoted peer learning and knowledge discovery and if students were able to discover and reuse knowledge efficiently. Results showed that participants felt MB was less difficult to complete all the scenarios than the Q&A system. When examining the time taken to complete the scenarios, we found participants completed the knowledge

discovery by keyword scenario faster with MB but completed the scenario for knowledge discovery using timeline faster with Q&A. We did not find significant differences between MB and Q&A in the time to complete the scenarios of knowledge reuse and discovery using patterns.

For the goal of facilitating completion of knowledge discovery and reuse tasks, we focused on success rate. We found that participants had a higher success rate for the knowledge discovery using keywords when using the MB system. Moreover, our results also showed that for the knowledge discovery using timeline scenario, participants had better success rate with Q&A. We also found that participants had user errors when interacting with the knowledge timeline filter in MB. Results also show that more than half of the participants integrated patterns while completing the knowledge discovery and reuse using patterns scenarios. Participants liked the idea of design patterns and found them effective at providing the desired information and useful.

For the goal of improving the user experience for browsing Q&A discussion threads, we found that the use of color-coding to denote popularity and node sizes to demonstrate answer count were easy to understand. Participants also felt that performing keyword and timeline searches were both useful and effective at providing the desired information. Participants also liked the integration of design patterns and the pattern browser, and the use of visual representations of the discussions. However, participants disliked the learning curve in using MB.

Our results also showed that participants not only found MB to be easy to use, but they agree with using it as part of their courses. For usage behavior, we presented classifications of participants based on their usage behavior for accessing details of discussions and browsing discussions using either the knowledge graph or the discussion listing view. Our results also

showed that undergraduate participants were more likely to browse discussions using the knowledge graph while participants with more years of college experience tended to prefer the discussion listing view.

# 6.0    FOLLOW UP DESIGN PATTERN PILOT STUDY


After implementation and evaluation of our innovative solution, we performed a follow up pilot study to further assess whether the integration of design patterns can contribute to peer-learning. The focus of this pilot study was on the integration of design patterns into the knowledge acquisition and reuse stages. Our user study described in Chapter 5 aimed to examine the ease of use of the MicroBrowser application and to assess that with the integration of design patterns, previous hard to complete educational tasks become easier to do. Results from our user study suggested that our design of MB was easy to use. Results also suggested that participants were able to not only integrate design patterns into the knowledge discovery and reuse process but they were also able to generate new knowledge in the form of design patterns and by creating associations between design patterns and discussion threads.

A limitation of the user study presented in Chapter 5 is that participants were not able to share the design patterns created by participants with each other. In our follow up pilot study, we did not want to restrict the available design patterns to those initially defined in the system but instead allow the knowledge base to evolve as new discussions are added and adapt as new design patterns are created and reused by participants. This design pattern creation and reuse process would allow the system to continuously adapt to its context and evolve with new knowledge created in the form of design patterns and their association to discussions.

We believe that the reuse of student-generated patterns enhances student learning because it might allow students to participate in the knowledge creation process and at the same time help other students to discover and create new knowledge. Thus, the goal of our follow up pilot study is to determine if participants make use of student generated design patterns.

The variables of interest in our pilot study consisted on the number of connections created between student-generated design patterns and discussions nodes, and the time it takes a participant to identify and associate a design pattern to a discussion.

## 6.1 METHODOLOGY

### 6.1.1 Apparatus

The apparatus was the same as in the user study described in Chapter 5, Section 1.1.1.

### 6.1.2 Outcomes

We focus on several outcomes for the pilot study: (1) the amount of time participants spent working on each task, (2) the amount of time participants took to identify an appropriate design pattern, (3) the number of design patterns participants used when completing a task, and (4) the count and percentage of student-generated patterns used.

### 6.1.3 Participants

We performed a controlled pilot study. Requirements to participate were to be college students in a field from a STEM major and being 18 years old or older.

### 6.1.4 Scenario

To simulate the scenario that participants were in a classroom setting, we provided participants with an introductory text about XML and its integration with Java. The material content described the structure for an XML document, the benefits of using XML to exchange messages between systems and two XML parsers (DOM and SAX) for processing XML messages in Java. We asked participants to read the material before starting the tasks. Below we show the introductory text:

---

*An Introduction to XML and Java*
*Extensible Markup Language (XML) is a markup language that defines a set of rules for encoding documents in a format which is both human-readable and machine-readable. A sample XML document looks like the following:*

```
<document>
  <chapter name="chapter01">
    <section name="Introduction">
      This is the text of the introduction of chapter 01
    </section>
    <section name="anothersection">
      This is another section in the same chapter
    </section>
  </chapter>
</document>
```

*The Java programming language contains several methods for processing and writing XML. By using these methods, a Java program can extract information from messages received by other programs and be able to read the contents and perform activities with the information received. A Java program can also generate an XML message to send information to other programs or to save the information as an XML file in the file system.*
*To read XML files in Java, you use an XML parser. There are two common methods: DOMParser and SAXParser. A DOMParser loads the entire XML file into memory allowing the Java program to easily access its elements and attributes. A SAXParser instead generates events as an XML file is being read and allowing the Java program to only use the elements and attributes it is interested on. A DOMParser is slow and consumes a lot of memory while a SAXParser is faster and better for large XML files.*

---

> **Patterns**
> *Patterns are solutions to problems that occur repeatedly in a given context. By describing a recurring problem in a general form, the solution to the problem can be reuse to solve similar problems that can appear under different situations. In a Question & Answer (Q&A) setting, we can assume that each question is a problem and the answer given is a solution. If the solution to the problem is to follow a particular step or process, then if we find another question that is similar but for a different topic, we can answer the question by referencing the appropriate pattern. In a general form, a pattern is composed of a name, a problem and a solution.*

### 6.1.5   Tasks

A set of tasks that aim to encourage participants to complete them by making use of a design pattern were defined. Four tasks were defined for the pilot study. Task 1 focused on knowledge creation by asking participant to create a new design pattern. Task 2 and Task 3 focused on knowledge discovery by asking participants to identify a design pattern from the list of available design patterns. The difference between Task 2 and Task 3 was that for Task 2 participants were not able to create new design patterns. Task 4 focused on using design patterns to help with answering a given question. For each task, we logged usage activity such as mouse clicks and system events, and start and end times.

### 6.1.6   Data

For the pilot study we reused the same data used during the user study, described in Chapter 5 Section 5.1.2, but expanded the list of available patterns by including all the student recommended patterns listed in Table 7. In total, there were 15 patterns, out of which 11 were student-generated in the user study.

## 6.2 RESULTS

### 6.2.1 Participants

A total of 5 participants (2 female, 40%) were enrolled in the follow up pilot study. Participants were students at the University of Pittsburgh and consisted of 3 Computer Science majors, 1 Information Science major and 1 Telecommunications major. Two (2) participants were graduate students and the mean years of college experience was 4.5 ($SD = 3.74$) years.

### 6.2.2 Can participants create design patterns with ease?

Our results showed that all participants (100%) completed Task 1 by creating a design pattern. Results also show that participants took an average of 220.8 seconds ($SD = 81.7$) to complete Task 1. In addition, participants took an average of 209.0 seconds ($SD = 45.7$) creating a design pattern. Table 16 summarizes the list of patterns that participants created during the pilot study.

**Table 16.** Patterns created by participants during follow-up session

| Pattern name | Pattern description | Pattern solution |
|---|---|---|
| Interview | You need to obtain information about a problem or functionality but there is no documentation available | Contact a person that has worked on the problem or functionality in the past to obtain more details |
| Reference Documentation | Maximize Productivity | Use validated reference |

| | | documentation |
|---|---|---|
| Testing | If you are trying to verify what you have implemented is what you intended | You should have a mean to validate your code/script by testing it |
| File Tracing | You are trying to capture activities that occurred while users interacted with your code or application and reviewing them after they closed the application | can capture the activities users performed in an archive in the file system to be reviewed after the application is closed |
| How to? | The pattern is defined to answer simple questions regarding HOW TO ...? | First find possible solution alternatives to the question and then explain step by step with the best solution chosen |

Based on the data, results suggest that the process of creating design patterns is easy to perform using the MB system and that participants can define patterns on their own. We asked participants how easy they felt it was to create a new design pattern using the interface. On a 5-point Likert-scale (1-Very easy, 5-Very hard) for ease of completing the task, the median response was a 1 ($IQR = 0$). Quality of the patterns created by participants was not evaluated in the pilot study. However, we believe there is an opportunity to further improve on this research by providing a pattern promotion process where patterns that are deemed of good quality and reusability become part of the pattern language.

One observation we noted in the pilot study was that most of the participants were unsure about how to articulate the pattern answer abstractly. Participants wanted to provide an answer to a problem and needed to get clarified that the solution should not apply to a specific problem but be expressed in a general form. Scaffolding techniques were used by (Muller, 2005) and (Ginat, 2009) to help students understand the use of design patterns. Similarly, we referenced the expert-generated answers to give participants examples of how to present the answers abstractly.

### 6.2.3 Can participants use patterns created by other students?

During the user study we performed earlier, participants were provided with initial design patterns and were asked to either use them or create their own. For this follow up pilot study, the goal was to investigate whether participants use not only expert-generated design patterns but also those created by other students. Task 2 and Task 3 asked students to find a popular discussion for the topic of XML.

Participants spent an average of 313 seconds ($SD$ = 193.0) and an average of 179.2 seconds ($SD$ = 96.4) to complete Task 2 and Task 3, respectively. Out of the total time spent on a task, participants spend an average of 138.4 seconds ($SD$ = 180.8) to select a pattern for Task 2 and an average of 59 seconds ($SD$ = 40.9) for Task 3.

For the activity of selecting a pattern, participants visited on average 8.0 patterns ($SD$ = 10.4) for Task 2 and 4.4 patterns ($SD$ = 3.3) for Task 3. By evaluating the data for the type of patterns visited, we found that 4 (80%) participants visited the details of a student generated

pattern in Task 2, while 4 (80%) participants did the same in Task 3. Regarding the actual patterns selected, 2 (40%) of the participants selected a student generated pattern for Task 2 and 2 (40%) of the participants did the same for Task 3. Figure 37 shows the pattern use frequencies for both tasks.

For the activity of finding an appropriate pattern, Task 3 had the additional option that if none of the existing patterns were helpful, participants could create their own pattern. Data shows that 1 (20%) of the participants created their own pattern and selected it for the pattern association. We also noted that the Educator pattern was the most commonly used pattern in Task 2, while for Task 3 participants reused different patterns. Results on the 5-point Likert-scale (1-very ease, 5-very hard) for ease of completing the task, the median response was a 2 (IQR=1) for Task 2 and a median response of 2 (IQR = 0) for Task 3.



(a) Task 2 pattern used                    (b) Task 3 pattern used
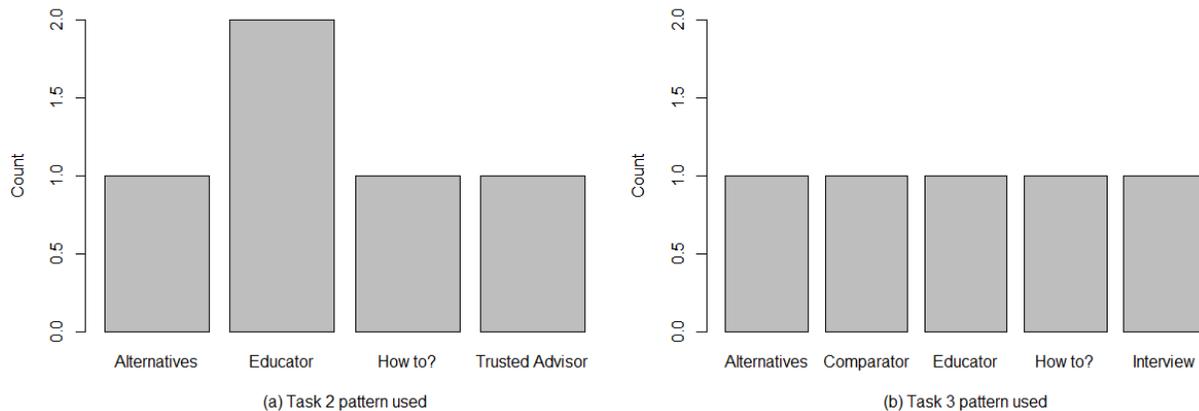
**Figure 37.** Frequency of patterns used when completing Tasks 2 and 3

Our goal for Tasks 2 and 3 was to investigate whether participants would reuse patterns created by other students. From the data, we found that participants reused the patterns created by other students in three forms. First, participants viewed the contents of the patterns for

153

consideration as a possible candidate to associate to a discussion. Second, participants selected a student generated pattern to associate to a discussion. And third, participants created their own patterns for association. These results suggest that pattern reuse occurs for student generated patterns. However, our data also shows that while participants viewed the details of a student-generated pattern, they did not select them. A possible explanation for this might originate in how specific a design pattern is. That is, if a design pattern specifically addresses discussions about File Tracing, then it would be hard for participants to reuse the pattern with questions that are not about File Tracing. The expert-generated patterns had a broader reach (Educator, Comparator, etc.) which facilitated its reuse across a more diverse set of questions.

### 6.2.4   Students benefit from access to patterns when answering discussions

For this goal, we wanted to know how participants benefit from using patterns while answering questions. We asked participants to answer a question but before doing so, to refer to a design pattern for guidance. For the use of design patterns when answering questions, 4 (80%) of the participants reported referencing an existing pattern to answer the question. Reviewing the data, results showed that only 3 (60%) participants associated a pattern to a discussion following the submission of an answer. From the patterns used, 2 (33%) of them were student generated. Participants visited on average 7.5 patterns ($SD = 3.317$) while completing Task 4. Among the participants that referenced a pattern, results in Table 17 show that the pattern that participants reviewed the most was the *Alternatives* followed by the *Reference Documentation* patterns. Students previously created both of these patterns. From the entire list of patterns, the data

154

shows that nine out of fourteen (64%) patterns reviewed were previously created by students. None of the participants created their own patterns as part of completing Task 4.

**Table 17.** Frequency of patterns explored

| Pattern | Frequency | Student generated |
|---|---|---|
| Alternatives | 4 | * |
| Reference Documentation | 3 | * |
| Educator | 3 | |
| Assignment / Homework Questions | 2 | * |
| Tutorials | 2 | * |
| Testing | 2 | * |
| Well answered / Well Asked | 2 | * |
| Trusted Advisor | 2 | |
| Summary | 1 | * |
| Reference Provider | 1 | |
| Debug | 1 | * |
| Comparator | 1 | |
| Previously answered | 1 | * |
| Reference Provider | 1 | |

Participants spent on average 233.6 seconds ($SD = 66.7$) to complete Task 4. For the activity of creating an answer, participants spent on average 41.7 seconds ($SD = 30.0$). On a 5-point Likert-scale (1-very ease, 5-very hard), the median response to how easy participants felt completing this task was a 2 ($IQR = 1$).

From the results of Task 4, we observe that even though participants could not tell the difference between student and non-student generated design patterns, 2 participants used student-generated patterns. It is important to note that even though only 2 participants used student-generate patterns, results show that the other 3 participants used expert-generated design patterns suggesting the reuse of design patterns by all participants.

## 6.3     LIMITATIONS

The goal of this pilot study was to understand how participants' learning behavior benefits from using student-generated design patterns. To assess this goal, we setup a controlled pilot study were participants were asked to complete a set a tasks. The use of a controlled pilot study introduced some limitations. The sample size was small at 5 participants. The reasoning behind the small number of participants enrolled in the pilot study was because it was meant to serve to determine the feasibility of the experiment and to plan for a larger user study. The pilot study was not performed on a classroom setting. This fact prevented us from considering factors such as participants' interaction, distractions and other situations that could have influenced the performance of participants with the system. Another limitation of the study is the degree at which participants were immersed into a learning mode. In the pilot study, we asked participants to review a small learning unit rather than asking them to reference lengthier classroom material, for example, questions based on a particular lecture.

## 6.4     CONCLUSION

The goal of this pilot study was to understand how participants' learning behavior benefits from student-generated design patterns. To analyze this goal, we defined three main scenarios. The first scenario explored whether participants could create design patterns with ease. Our results show that participants were able to create design patterns with ease. However, we noted that while the process of using MB to create the design pattern was easy, participants had difficulty

stating the solutions for a pattern abstractly that could facilitate its reuse. A recommendation could be to include more samples in the introductory text.

For the goal of reusing student-generated patterns, our results show that participants explored student-generated patterns when looking to associating a pattern to a discussion. While we expected more than half of participants to select a student-generated pattern, the fact that at least some of the participants did select one was encouraging. We believe that student-generated design patterns might be too specific to particular scenarios limiting their applicability to other scenarios.

Finally, for the goal of referencing student-generated patterns when answering questions, our results showed that while the majority of participants viewed the details of student-generated patterns, less than half ultimately selected a student-generated pattern. Similar to the prior goal, we believe that the choice of pattern selection by participants might be affected by how abstract patterns were.

Experiences with the pilot study can help us in the planning to conduct a larger user study. Some of the issues identified in the pilot study were the need to examine the accuracy and quality of the design patterns in addition to their use. Future work can explore the implementation of validation rubrics to quantify the goodness of a pattern. We would also like to explore the effects from utilizing different ratios of student-generated design patterns and expert- and student-generated design patterns.

# 7.0    CONCLUSION AND FUTURE WORK

In this work we propose an innovative approach that integrates design patterns with an interactive knowledge graph to facilitate knowledge reuse and knowledge discovery in large knowledge bases. We designed and implemented a streamlined knowledge-graph navigation system augmented with 2D discussion visualizations and pattern based expertise-sharing interfaces in fields of STEM majors called MicroBrowser. The MicroBrowser system provides a set of pattern-based expertise sharing interfaces to allow both learners and instructors to refine, highlight, reuse, and share knowledge. MicroBrowser also allows learners to browse and navigate important discussions based on topic similarity encoded by node proximity in a knowledge graph. The description of our approach is presented in Chapter 4.

Inspired by the reuse capabilities of design patterns and the effectiveness of Q&A systems as knowledge repositories, we incorporated design patterns in the proposed approach as means to provide solutions to recurring questions in educational domains. One goal of this work was to determine if the proposed novel approach facilitated knowledge discovery and knowledge reuse by the use of design patters as insightful reusable knowledge nuggets. Results from the user study and the pilot study suggest that students were able to use design patterns to complete tasks that required the identification of discussions based on specific criteria such as how the question were answered, and how to improve a question for reusability, among others. In

addition, participants were able to reuse design patterns by assigning an appropriate design patterns to a specified discussion. Results of our empirical evaluations also suggest that task difficulty was perceived to be lower using MB when compared to state-of-the-art Q&A system for all the knowledge discovery and reuse scenarios evaluated. For the scenario of knowledge discovery using keywords, not only was success rate higher but also participants took less time to complete the scenario tasks when using MB. Moreover, results also suggest that participants found the pattern-based expertise-sharing interface effective and easy to understand. Results of the success rate for the scenarios of using patterns for knowledge discovery and reuse show that even though it was a new concept, the majority of the participants integrated the use of patterns as part of the solution. Results showed that participants were able to effectively integrate the use of design patterns for knowledge discovery, generation and reuse and are consistent with prior work. Lin et al. (Lin & Landay, 2008) showed how designers reused patterns to build web UIs for distinct interfaces. In addition, Carle et al. demonstrated how novice instructors learned from expert-generated patterns to design course content (Carle et al., 2007). Muller et al. (Muller, Ginat, & Haberman, 2007) showed that teaching algorithmic patterns as part of an introductory Computer Science course helped students to construct solutions to problems by being able to decompose the problem into sub tasks and to associate algorithmic patterns to solve the sub tasks. Wilkinson et al. described an approach of using computational patterns to help students solve parallel programming tasks by discovering and reusing computational patterns (Wilkinson, Villalobos, & Ferner, 2013).

Prior work has shown that the use of visualization techniques take advantage of humans' visual exploration capabilities allowing students to effectively discover knowledge from discussion threads in the form of content, attributes, relationships, and time information (Heer et

al., 2012). Therefore, our proposed system also integrated the use of visualization techniques to facilitate knowledge discovery in large knowledge bases. Our system represented knowledge using an interactive knowledge graph with nodes representing discussion threads or design patterns, and edges representing relationships. To address the challenge of finding popular discussions, we measured popularity as the number of page visits and represented the popularity using a color gradient scheme with darker color hues representing more popular items. Similar to prior work, we encoded discussion popularity using color gradient schemes (Xia, Li, Wang, & Gu, 2007)(Wanner et al., 2011). To quickly identify the current state (answered or unanswered) of a discussion, we represented discussion nodes using different shapes. We also made us of shapes in combination with the pattern name to represent design patterns on the knowledge graph. Prior work have also used shapes to encode information in visualizations such as semantic profiles of documents (DeCamp et al., 2005) and to denote blog entries, comments and number of words (Indratmo & Gutwin, 2008). Participants agreed that the use of shapes to differentiate between answered and unanswered discussions, the use of color-coding to demonstrate popularity and the use of different node sizes to represent answer counts were easy to understand and effective at providing the desired information.

Topic similarity was used in the MicroBrowser system to measure the relationship between two discussions. To represent the strength of the relationship, we encoded the strength as the intensity of the stroke color of the edge between two nodes. To address the challenge of overcrowding the knowledge graph with too many edges, we introduced a similarity slider to allow participants to progressively display additional connections based on the similarity strength. Prior studies have shown that in educational settings, peak demand for knowledge discovery is near deadlines (Warren et al., 2014). Therefore, we also introduced the use of a

knowledge timeline to allow students to control the visibility of discussions based on their temporal attributes.

Limitations of our user study included the small sample size, the fact that the user study was performed in a controlled environment might hide potential interactions among participants that could occur in a classroom setting, the limited time participants had to learn about the advantages and use of design patterns, and the learning curve with learning a system that was different from traditional available Q&A systems. Even though our user study and pilot study showed that MicroBrowser was shown to be effective at facilitating knowledge discovery and reuse, future work will improve features that participants recommended. As an example, future improvement for the MicroBrowser system should focus on the ability to sort discussions based on different attributes and improvement to the knowledge timeline to reduce the number of errors from selecting date ranges. Feedback from participants also addressed the need to use more distinct shapes to represent the different states. Even though participants were able to use and reuse design patterns, we noted the need to provide more scaffolding to assist participants to describe the solution to a pattern in a reusable form. We believe that with longer exposure to the system and design patterns, participants will be able to acquire more expertise and familiarity using design patterns.

The integration of design patterns in knowledge discovery and reuse is a key innovation in the MicroBrowser system. Overall, the MicroBrowser system achieves this with an interactive knowledge visualization and exploration system. The most liked feature in the system was the integration of design patterns and the pattern browser followed by the visual representation of discussions. Our system is not limited to the fields of Computer Science and Information Sciences in STEM majors but can also be beneficial to other majors.

## 7.1    FUTURE WORK

As discussed in the conclusion, the MicroBrowser system was effective at facilitating knowledge discovery and reuse by combining design patterns and visualization techniques.    Future work should focus on extending the user study and pilot study not only to a larger population but also for a longer period of time.  In particular, we would like to be able to evaluate MB in multiple classroom settings to further explore how MB enhances peer learning and to further investigate how students use patterns to enhance their learning.  Performing the user study in classroom settings can also allow us to measure acquired knowledge by comparing for example pre- and post-assessment evaluations.  By extending the study over a longer period, we could also investigate the evolution of the knowledge base and interaction patterns between students and MB.

Future work should also explore the ability to automatically extract patterns from existing discussion threads based on discussion characteristics such as popularity, age, and number of responses among others.  For example, if a discussion node is very popular then students must find it useful.  Furthermore, we would like to assess if it is possible to automatically create a design pattern using the title of the question as the name of the pattern, the text as the description and the accepted answer as the solution based on the popularity of a discussion.

In addition, we would like further the process of generating and incorporating student generated patterns into a pattern language.  We would also like to explore techniques for

validating student-generated patterns and techniques for validating the quality of student-generated patters.  For example, in the user study, participants were able to create their own patterns and add relationships to other patterns or discussions.  However, as more design patterns are created, the likelihood of bogus or invalid patterns increase.  Even if the pattern was valid, it is possible the association to other patterns or discussions is not.  To prevent this, we would like to devise methods that help not only to manage the process of how new design patterns become members of the pattern language but also to test the quality and validity of their associations.

Our work is not limited to the fields of Computer Science or Information Science education but can be extended to other fields and STEM majors as well.  One approach we have considered is to make MicroBrowser pluggable into Learning Management Systems or Learning Content Management Systems, such as Moodle, Blackboard, among others.

# APPENDIX A

## USER STUDY TASKS

### A.1    EXPERIMENTAL TASKS SET - MICROBROWSER

#### A.1.1  SCENARIO: Training

**PRACTICE TASK 1:** Can you find a discussion that covers the topic of **unix**.  Can you determine if the question has been answered?

**ANSWER (Discussion ID)**: _____

**Has it been answered?**  Yes  No

**PRACTICE TASK 2**: A pattern not only allows you to find additional information about discussions but it also helps you discover new information such as examples of how a discussion has been answered.  Can you locate a discussion that has been associated to a pattern? After you

find the discussion, open the details of the pattern and locate another discussion that has been associated to the pattern.

**ANSWER (Discussion ID)**: _____

**Discussion associated to the same pattern:_____**

## A.1.2  SCENARIO: Finding discussion based on topic searches

**TASK 1:**       During the last class, the instructor discussed topics on Java testing frameworks, in particular the use of JUnit.  Can you find one discussion related to the concept of **JUnit**?

**ANSWER (Discussion ID)**: _____

---

**How difficult was to complete the task using the application?**

| Very difficult | Difficult | Neutral | Easy | Very easy |

**How likely is that an instructor would have asked you to complete a similar task?**

| Very unlikely | Unlikely | Neutral | Likely | Very likely |

---

**TASK 2:**       Assume you are trying to come up with discussion points for your next class which covers the use of XML.  Can you find two discussions related to the topic of **XML** that students have viewed the most?

**ANSWER (Discussion ID)**: _____   _____

**How difficult was to complete the task using the application?**

| Very difficult | Difficult | Neutral | Easy | Very easy |

**How likely is that an instructor would have asked you to complete a similar task?**
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

**TASK 3:**     As part of a homework assignment, the instructor has asked you to provide an answer to a question about the topic **networking**.  To learn more about the topic, you decide to read similar questions. To complete this task, first find an unanswered question (a question without an accepted answer) about the topic **networking** and then find two related questions.

ANSWER (Discussion ID): _____     _____

**How difficult was to complete the task using the application?**
| Very difficult | Difficult | Neutral | Easy | Very easy |

**How likely is that an instructor would have asked you to complete a similar task?**
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

## A.1.3  SCENARIO: Finding time based information

**TASK 4:**     Supposed you had a personal emergency that prevented you from attending class during the first week of the year and now you want to know what topics were covered when you were out.  Can you find activity that occurred between **Monday September 22$^{nd}$** and **Friday September 26$^{th}$, 2014**? (Pick any 2 discussions you identify)

ANSWER (Discussion ID): _____     _____

**How difficult was to complete the task using the application?**
| Very difficult | Difficult | Neutral | Easy | Very easy |

**How likely are you to have performed this task in a classroom setting?**
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

**TASK 5:**     During the month of **July 2014**, the instructor covered Java networking and the use of **sockets** for writing Internet applications.   The instructor suggested reviewing discussions about this topic.  Can you find some questions related to using **sockets** during the month of July 2014?  (Pick any 2 discussions you identify)

ANSWER (Discussion ID): _____          _____

| How difficult was to complete the task using the application? | | | | |
|---|---|---|---|---|
| Very difficult | Difficult | Neutral | Easy | Very easy |

| How likely is that an instructor would have asked you to complete a similar task? | | | | |
|---|---|---|---|---|
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

**TASK 6:**     (NOTE: To complete this task you might need to change the dates) Suppose you have been given an assignment to help answering classmates' questions. To prepare for this assignment, you decide to find the most recent question which does not currently have an **accepted** answer. Can you find it?

ANSWER (Discussion ID): _____

| How difficult was to complete the task using the application? | | | | |
|---|---|---|---|---|
| Very difficult | Difficult | Neutral | Easy | Very easy |

| How likely is that an instructor would have asked you to complete a similar task? | | | | |
|---|---|---|---|---|
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

## A.1.4  SCENARIO: Using patterns to improve the quality of answers

**TASK 7:** Suppose the instructor asked you to collaborate with your classmates by answering questions posted by each other requesting recommendations or advice.  To gain experience for how to complete the assignment, you decide to find some examples where the questions have been answered by providing either advice or recommendations.  Can you find two examples?

**ANSWER (Discussion ID)**: _____        _____

| How difficult was to complete the task using the application? | | | | |
|---|---|---|---|---|
| Very difficult | Difficult | Neutral | Easy | Very easy |
| **How likely is that an instructor would have asked you to complete a similar task?** | | | | |
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

**TASK 8:** Browsing discussions about Regular Expressions (keyword: **regex**), you believe discussion **Using Regex to generate Strings rather than match them** might helpful for students looking for examples of discussions that <u>provide references</u>.  Can you find the discussion and improve it so that it can be useful for other students looking for examples on answering questions using external references? (hint: how can you make it easier for other students looking for these type of examples to find it)

**ANSWER (What did you do)**: _____

_____

| Rating: | | | | |
|---|---|---|---|---|
| Very difficult | Difficult | Neutral | Easy | Very easy |
| **How likely is that an instructor would have asked you to complete a similar task?** | | | | |
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

**TASK 9:** Can you find an unanswered question about creating a "**welcome page** in **struts"** and submit an answer for it: **To create a welcome page, I recommend you to define a Welcome File List and then redirect to the default action**.  After submitting the answer, can you think of a way to allow other students to re-use how you presented your answer to the question?

**ANSWER (What did you do)**: _____

_____

| How difficult was to complete the task using the application? | | | | |
|---|---|---|---|---|
| Very difficult | Difficult | Neutral | Easy | Very easy |
| **How likely is that an instructor would have asked you to complete a similar task?** | | | | |

| Very unlikely | Unlikely | Neutral | Likely | Very likely |
|---|---|---|---|---|

## A.1.5 SCENARIO: Using Pattern information to perform knowledge discovery

**TASK 10:** Suppose you are trying to reply to a question posted by a student by giving a reference to a tool you are familiar with but you are not sure how best to state your answer.  To give you some guidance, can you find two examples of how other students have answered question by referencing other tools as part of their answer.

**ANSWER (What did you do)**: _____

_____

**ANSWER (Discussion ID)**: _____      _____

| **How difficult was to complete the task using the application?** | | | | |
|---|---|---|---|---|
| Very difficult | Difficult | Neutral | Easy | Very easy |
| **How likely is that an instructor would have asked you to complete a similar task?** | | | | |
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

**TASK 11:** The instructor has given you an assignment to perform a comparison between two components.  The instructor suggested referring to question **Differences between Hashmap and Hashtable (keyword: hashmap)** as an example of how you could complete the assignment but recommended you should look at other examples as well.  Can you find another sample question which compares between two or more components?

**ANSWER (What did you do)**: _____

_____

**ANSWER (Discussion ID)**: _____

| **How difficult was to complete the task using the application?** | | | | |
|---|---|---|---|---|
| Very difficult | Difficult | Neutral | Easy | Very easy |

| How likely is that an instructor would have asked you to complete a similar task? | | | | |
|---|---|---|---|---|
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

**TASK 12:** Your homework for this week is to answer a question from the list of unanswered ones. However, you are afraid of wording your answer incorrectly. Can you find two examples for recommended ways to structure your answers?

**ANSWER (What did you do)**: _____

_____

**ANSWER (Discussion ID)**: _____ _____

| How difficult was to complete the task using the application? | | | | |
|---|---|---|---|---|
| Very difficult | Difficult | Neutral | Easy | Very easy |
| **How likely is that an instructor would have asked you to complete a similar task?** | | | | |
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

## A.2    EXPERIMENTAL TASKS SET – MICROPROBE

### A.2.1   SCENARIO: Finding discussion based on topic searches

**TASK 1:** During the last class, the instructor discussed topics on Java scripting frameworks, in particular the use of Groovy. Can you find one discussion related to the concept of **Groovy**?

**ANSWER (Discussion ID)**: _____

| How difficult was to complete the task using the application? | | | | |
|---|---|---|---|---|
| Very difficult | Difficult | Neutral | Easy | Very easy |

| How likely is that an instructor would have asked you to complete a similar task? | | | | |
|---|---|---|---|---|
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

**TASK 2:** Assume you are trying to come up with discussion points for your next class which covers the use of Android programming. Can you find two discussions related to the topic of **Android** that students have viewed the most?

**ANSWER (Discussion ID)**: _____        _____

| How difficult was to complete the task using the application? | | | | |
|---|---|---|---|---|
| Very difficult | Difficult | Neutral | Easy | Very easy |

| How likely is that an instructor would have asked you to complete a similar task? | | | | |
|---|---|---|---|---|
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

**TASK 3:** As part of a homework assignment, the instructor has asked you to provide an answer to a question about the topic **code coverage**. To learn more about the topic, you decide to read similar questions. To complete this task, first find an unanswered question (a question without an accepted answer) about the topic **code coverage** and then find two related questions.

ANSWER (Discussion ID): _____          _____

| How difficult was to complete the task using the application? | | | | |
|---|---|---|---|---|
| Very difficult | Difficult | Neutral | Easy | Very easy |
| | | | | |
| **How likely is that an instructor would have asked you to complete a similar task?** | | | | |
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

## A.2.2   SCENARIO: Finding time based information

**TASK 4:** Supposed you had a personal emergency that prevented you from attending class during the first week of the year and now you want to know what topics were covered when you were out. Can you find activity that occurred between **August 20$^{th}$ 2008** and **August 25$^{th}$ 2008**? (Pick any 2 discussions you identify)

ANSWER (Discussion ID): _____          _____

| How difficult was to complete the task using the application? | | | | |
|---|---|---|---|---|
| Very difficult | Difficult | Neutral | Easy | Very easy |
| | | | | |
| **How likely is that an instructor would have asked you to complete a similar task?** | | | | |
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

**TASK 5:** During the month of **August 2008**, the instructor covered the benefits of using a **template** for consistency and speed when writing Java Swing applications. The instructor suggested reviewing discussions about this topic. Can you find some questions related to using templates during the month of August 2008? (Pick any 2 discussions you identify)

**ANSWER (Discussion ID)**: _____        _____

| How difficult was to complete the task using the application? | | | | |
|---|---|---|---|---|
| Very difficult | Difficult | Neutral | Easy | Very easy |

| How likely is that an instructor would have asked you to complete a similar task? | | | | |
|---|---|---|---|---|
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

**TASK 6:** Suppose you have been given an assignment to help answering classmates' questions. To prepare for this assignment, you decide to find the most recent question which do not currently have an **accepted** answer. Can you find it?

**ANSWER (Discussion ID)**: _____

| How difficult was to complete the task using the application? | | | | |
|---|---|---|---|---|
| Very difficult | Difficult | Neutral | Easy | Very easy |

| How likely is that an instructor would have asked you to complete a similar task? | | | | |
|---|---|---|---|---|
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

## A.2.3   SCENARIO: Using Pattern information to perform knowledge discovery

**TASK 7:** Suppose the instructor asked you to collaborate with your classmates by answering questions posted by each other requesting recommendations or advice. To gain experience for how to complete the assignment, you decide to find some examples where the questions have been answered by referencing other tools as part of their answer. Can you find two examples?

**ANSWER (Discussion ID)**: _____        _____

| How difficult was to complete the task using the application? | | | | |
|---|---|---|---|---|
| Very difficult | Difficult | Neutral | Easy | Very easy |

**How likely is that an instructor would have asked you to complete a similar task?**
Very unlikely | Unlikely | Neutral | Likely | Very likely

**TASK 8:** Browsing discussions about Java testing frameworks, you believe discussion `Junit vs TestNG` might helpful for students looking for examples of discussions that <u>provide a comparison between two applications</u>. Can you find the discussion and improve it so that it can be useful for other students looking for examples on answering questions using external references? (hint: how can you make it easier for other students looking for these type of examples to find it)

**ANSWER (What did you do)**: _____

_____

| How difficult was to complete the task using the application? | | | | |
|---|---|---|---|---|
| Very difficult | Difficult | Neutral | Easy | Very easy |

**How likely is that an instructor would have asked you to complete a similar task?**
Very unlikely | Unlikely | Neutral | Likely | Very likely

**TASK 9:** Suppose you have been asked to answer a question from the list of unanswered ones. However, you are afraid of wording your answer incorrectly. Can you find two examples for recommended ways to structure your answers?

**ANSWER (What did you do)**: _____

_____

**ANSWER (Discussion ID)**: _____         _____

**How difficult was to complete the task using the application?**

| Very difficult | Difficult | Neutral | Easy | Very easy |

**How likely is that an instructor would have asked you to complete a similar task?**
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

## A.2.4   SCENARIO: Using patterns to improve the quality of answers

**TASK 10:**       Suppose the instructor noted you are an expert on writing SQL applications and has asked you to answer questions where users ask for recommendations or advice.  To gain experience for how to state your answers, you decide to find some examples, even if not related to SQL, where the questions have been answered by providing recommendations.  Can you find two examples?

**ANSWER (What did you do)**: _____

_____

**ANSWER (Discussion ID)**: _____          _____

**How difficult was to complete the task using the application?**
| Very difficult | Difficult | Neutral | Easy | Very easy |

**How likely is that an instructor would have asked you to complete a similar task?**
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

**TASK 11:**       Browsing discussions about Java AOP (keyword: **Java AOP**), you believe discussion **Where can I find a good tutorial about Java AOP that is proxy based?** to be very helpful for other students looking for examples of discussions that provide references.  Can you find the discussion and improve it so that it can be useful for other students looking for examples on answering questions using external references? (hint: how can you make it easier for others looking for these type of examples to find it)

**ANSWER (What did you do)**: _____

| How difficult was to complete the task using the application? | | | | |
| --- | --- | --- | --- | --- |
| Very difficult | Difficult | Neutral | Easy | Very easy |
| **How likely is that an instructor would have asked you to complete a similar task?** | | | | |
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

**TASK 12:** Find the discussion **Java File IO Compendium** and submit the following answer for it: **From a previous project I worked on, I found that the following tutorial provides a good example of file manipulation using Java: http://docs.oracle.com/javase/tutorial/essential/io.** Can you think of a way to allow other users to re-use how you presented your answer to the question?

**ANSWER (What did you do)**: _____

_____

| How difficult was to complete the task using the application? | | | | |
| --- | --- | --- | --- | --- |
| Very difficult | Difficult | Neutral | Easy | Very easy |
| **How likely is that an instructor would have asked you to complete a similar task?** | | | | |
| Very unlikely | Unlikely | Neutral | Likely | Very likely |

# APPENDIX B

## MICROBROWSER – QUESTIONNAIRE

## B.1 PART I: EXPERIENCE WITH DISCUSSION FORUMS OR BOARDS

1. Prior to this study, were you familiar with discussion forums?
   __ Yes          ___ No

2. Are you familiar with Question & Answers sites such as StackOverflow, Yahoo! Answers, etc?

   __ Yes          __ No

3. If you had experience with discussion forums before, indicate whether you had any past experience with any of the following:

|                                | Past Experience | |
|--------------------------------|:---:|:---:|
|                                | Yes | No |
| Browsed questions and answers  |     |    |

177

| | | |
|---|---|---|
| Created questions | | |
| Submitted an answer | | |
| Voted on answers | | |
| Voted on questions | | |
| Viewed their leaderboard | | |
| Created an account | | |

4. Have you used discussion forums that are part of courses you have taken?

___ Yes ___ No

| | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 5. If you have used discussion forums on any of your classes before, to what extent do you agree they were helpful for learning course content? | ○ | ○ | ○ | ○ | ○ |
| 6. If you have NOT used discussion forums for any of your classes before, to what extent do you agree that they could be helpful for learning course content? | ○ | ○ | ○ | ○ | ○ |

## B.2    PART II: USING MICROBROWSER

| | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 7. It was easy to find discussions on particular topics using the MicroBrowser application. | ○ | ○ | ○ | ○ | ○ |
| 8. It was easy to find discussions that were related to the topics of interest using the MicroBrowser application. | ○ | ○ | ○ | ○ | ○ |
| 9. It was easy to understand the concept of Design Patterns and how they are associated to discussions when using the MicroBrowser application. | ○ | ○ | ○ | ○ | ○ |
| 10. It was easy to find popular discussions using the MicroBrowser application. | ○ | ○ | ○ | ○ | ○ |
| 11. It was easy to find answered and unanswered questions using the MicroBrowser application. | ○ | ○ | ○ | ○ | ○ |

| | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 12. The use of design patterns made finding related discussions easier when using the MicroBrowser application. | ○ | ○ | ○ | ○ | ○ |
| 13. The use of different colors to demonstrate popularity was easy to understand when using the MicroBrowser application. | ○ | ○ | ○ | ○ | ○ |
| 14. Using different sizes to show answer counts was easy to understand when using the MicroBrowser application. | ○ | ○ | ○ | ○ | ○ |

| | | | | | |
|---|---|---|---|---|---|
| 15. Having pattern information can help me to learn how to write my questions when using a Q&A system. | ○ | ○ | ○ | ○ | ○ |
| 16. Having pattern information can help me to learn how to answer questions when using a Q&A system. | ○ | ○ | ○ | ○ | ○ |
| 17. Having pattern information can motivate me to answer questions on a Q&A system? | ○ | ○ | ○ | ○ | ○ |

18. Overall, rate how easy to use you found the MicroBrowser application.

Not easy
at all    1    2    3    4    5    6    7    8    9    10      Very easy
to use

19. Did you use the keyword search and filter feature of MicroBrowser?  If yes, please answer question #20 below.  If not, why did not you use this function?

20. I feel the "keyword search and filter feature" function in MicroBrowser is:

| | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|---|
| Useful | ○ | ○ | ○ | ○ | ○ |

| | | | | | |
|---|---|---|---|---|---|
| Easy to Understand | ○ | ○ | ○ | ○ | ○ |
| Easy to Learn | ○ | ○ | ○ | ○ | ○ |
| Easy to notice | ○ | ○ | ○ | ○ | ○ |
| Effective in providing the desired information | ○ | ○ | ○ | ○ | ○ |

21. Did you use the "start and end date filter" function of MicroBrowser?  If yes, please answer question #22 below.  If not, why did not you use this function?

22.  I feel the "start and end date filter"  function in MicroBrowser is:

| | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|---|
| Useful | ○ | ○ | ○ | ○ | ○ |
| Easy to Understand | ○ | ○ | ○ | ○ | ○ |
| Easy to Learn | ○ | ○ | ○ | ○ | ○ |
| Easy to notice | ○ | ○ | ○ | ○ | ○ |
| Effective in providing the desired information | ○ | ○ | ○ | ○ | ○ |

23. Did you use the "pattern associations to discussion threads" feature of MicroBrowser?  If yes, please answer question #24 below.  If not, why did not you use this function?

24. I feel the "pattern associations to discussion threads" function in MicroBrowser is:

| | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|---|
| Useful | ○ | ○ | ○ | ○ | ○ |
| Easy to Understand | ○ | ○ | ○ | ○ | ○ |
| Easy to Learn | ○ | ○ | ○ | ○ | ○ |
| Easy to notice | ○ | ○ | ○ | ○ | ○ |
| Effective in providing the desired information | ○ | ○ | ○ | ○ | ○ |

25. Did you use the "discussion details view" feature of MicroBrowser?  If yes, please answer question #26 below.  If not, why did not you use this function?

26. I feel the "discussion details view" function in MicroBrowser is:

| | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|---|
| Useful | ○ | ○ | ○ | ○ | ○ |
| Easy to Understand | ○ | ○ | ○ | ○ | ○ |
| Easy to Learn | ○ | ○ | ○ | ○ | ○ |
| Easy to notice | ○ | ○ | ○ | ○ | ○ |

| | | | | | |
|---|---|---|---|---|---|
| Effective in providing the desired information | ○ | ○ | ○ | ○ | ○ |

## B.3    PART III: COMPARING TRADITIONAL Q&A SITE WITH MICROBROWSER

| | **Q&A is easier** | | **Neutral** | | **MicroBrowser is easier** |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| 27. Finding questions | ○ | ○ | ○ | ○ | ○ |
| 28. Finding unanswered questions | ○ | ○ | ○ | ○ | ○ |
| 29. Finding answered questions | ○ | ○ | ○ | ○ | ○ |
| 30. Finding questions based on time | ○ | ○ | ○ | ○ | ○ |
| 31. Ease of use | ○ | ○ | ○ | ○ | ○ |

32. I would consider MicroBrowser to be an improvement to the traditional Q&A Site.

    Strongly Disagree        Disagree        Neutral        Agree        Strongly Agree

## B.4    PART IV: USING MICROBROWSER IN AN EDUCATION SETTING.

| | **Strongly Disagree** | **Disagree** | **Neutral** | **Agree** | **Strongly Agree** |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| 33. I would like to be able to use MicroBrowser in one of my courses. | ○ | ○ | ○ | ○ | ○ |
| 34. Using MicroBrowser in a course can help with knowledge sharing between classmates. | ○ | ○ | ○ | ○ | ○ |
| 35. Using MicroBrowser in a course can help with knowledge creation, for example, by creating new patterns, questions or answers. | ○ | ○ | ○ | ○ | ○ |
| 36. Using MicroBrowser in a course would motivate me to ask more questions. | ○ | ○ | ○ | ○ | ○ |
| 37. Using MicroBrowser in a course would motivate me to answer questions. | ○ | ○ | ○ | ○ | ○ |

## B.5    PART V: ADDITIONAL INFORMATION

38. What features in the MicroBrowser application did you like the most?

39. What features in the MicroBrowser application you did **NOT** like?

40. What new features could be added to improve the MicroBrowser application?

41. if you can recommend a new pattern to be added, what would you define it as:
    a. PATTERN NAME: _____
    b. PATTERN DESCRIPTION: _____
       _____
    c. PATTERN SOLUTION: _____
       _____

# BIBLIOGRAPHY

Agichtein, E. (2008). Finding the Right Facts in the Crowd : Factoid Question Answering over Social Media Categories and Subject Descriptors. *Search*, 467–476.

Alexander, C. (1979). *The Timeless Way of Building*. New York: Oxford University Press.

Alexander, C. (1999). The Origins of Pattern Theory: The Future of the Theory, and the Generation of a Living World. *IEEE Software*, *16*(5), 71–82. doi:10.1109/52.795104

Alexander, C., Ishikawa, S., & Silverstein, M. (1977). *A Pattern Language: Towns, Buildings, Construction*. Oxford University Press.

Answerbag. (n.d.). Retrieved from http://www.answerbag.com

Ardichvili, A., Page, V., & Wentling, T. (2003). Motivation and barriers to participation in virtual knowledge-sharing communities of practice. *Journal of Knowledge Management*, *7*(1), 64–77. doi:10.1108/13673270310463626

Aritajati, C., & Narayanan, N. H. (2013). Facilitating Students ' Collaboration and Learning in a Question and Answer System. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion*, 101–105. doi:10.1145/2441955.2441983

AskVille. (n.d.). Retrieved January 1, 2012, from http://askville.amazon.com

Astrachan, O., Mitchener, G., Berry, G., & Cox, L. (1998). Design patterns: an essential component of CS curricula. *SIGCSE '98: Proceedings of the Twenty-Ninth SIGCSE Technical Symposium on Computer Science Education*, 153–160. doi:http://doi.acm.org/10.1145/273133.273182

Barnes, J., & Hut, P. (1986). A hierarchical O (N log N) force-calculation algorithm. *Nature*, *324*, 446 – 449. doi:10.1038/324446a0

Bergin, J. (2012). *Pedagogical Patterns: Advice for Educators*. CreateSpace Independent Publishing Platform; Compact edition (August 21, 2012).

Bratitsis, T., & Dimitracopoulou, A. (2006). Monitoring and Analyzing Group Interactions in Asynchronous Discussions with the DIAS System. In *Groupware: Design, Implementation, and Use* (Vol. 4154, pp. 54–61). doi:10.1007/11853862_5

Brown, R. A., Adams, J. C., Ferner, C., Shoop, E., & Wilkinson, A. B. (2014). Teaching parallel design patterns to undergraduates in computer science. In *Proceedings of the 45th ACM*

*technical symposium on Computer science education - SIGCSE '14* (pp. 547–548). New York, New York, USA: ACM Press. doi:10.1145/2538862.2538875

Cai, Q., & Sheth, N. (2003). Visualizing MeSH Dataset using Radial Tree Layout.

Carle, A., Clancy, M., & Canny, J. (2007). Working with pedagogical patterns in PACT. *ACM SIGCSE Bulletin*. doi:10.1145/1227504.1227392

Carmel, D., Roitman, H., & Yom-Tov, E. (2012). On the Relationship between Novelty and Popularity of User-Generated Content. *ACM Transactions on Intelligent Systems and Technology*, *3*(4), 1–19. doi:10.1145/2337542.2337554

Chang, S.-K., Colace, F., Santo, M. De, Zegarra, E., & Qie, Y. (2010). MAMA : A Novel Approach to Ontology Mapping. In *Proceedings of the 16th International Conference on Distributed Multimedia Systems, DMS 2010*.

Chang, S.-K., Zegarra, E., Shih, C. C., Peng, T.-C., Colace, F., & De Santo, M. (2010). Ontological Filters for Slow Intelligence Systems. In *Proceedings of the 16th International Conference on Distributed Multimedia Systems, DMS 2010* (pp. 33–40). Retrieved from http://dblp.uni-trier.de/db/conf/dms/dms2010.html#ChangZSPCS10

Chizmar, J. F., & Williams, D. B. (1996). Altering Time and Space through Network Technologies to Enhance Learning. *Cause/Effect*, *19*, 14–21.

Chuang, J., Ramage, D., Manning, C. D., & Heer, J. (2012). Interpretation and Trust : Designing Model-Driven Visualizations for Text Analysis, 443–452.

Coetzee, D., Fox, A., & Hearst, M. A. (2014). Should Your MOOC Forum Use a Reputation System ?, *0*.

Cook, S., Boness, K., Bartsch, M., Harrison, R., & Wernick, P. (2005). A pattern language for goals. Retrieved from http://centaur.reading.ac.uk/6106/

Crouch, C. H., & Mazur, E. (2001). Peer Instruction: Ten years of experience and results. *American Journal of Physics*, *69*(9), 970. doi:10.1119/1.1374249

De Choudhury, M., Sundaram, H., John, A., & Seligmann, D. D. (2009). What Makes Conversations Interesting ? Themes , Participants and Consequences of Conversations in Online Social Media, 331–340.

Dearman, D., & Truong, K. N. (2010). Why Users of Yahoo ! Answers Do Not Answer Questions. *Science*, 2008–2011.

DeCamp, P., Frid-Jimenez, A., Guiness, J., & Roy, D. (2005). Gist icons: Seeing meaning in large bodies of literature. In *Proceedings of IEEE Symposium on Information Visualization, Poster Session*. IEEE CS Press. Retrieved from http://media.mit.edu/cogmac/publications/IEEEIcons.pdf

Dellarocas, C. (2010). Designing Reputation Systems for the Social Web. *Boston University School of Management Research Paper Series*, *18*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1624697

Deng, J., Kemp, E., & Todd, E. G. (2005). Managing UI pattern collections. *Proceedings of the 6th ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction Making CHI Natural - CHINZ '05*, 31–38. doi:10.1145/1073943.1073951

Deterding, S., Dixon, D., & Khaled, R. (2011). Gamification: Toward a Definition. In *CHI 2011 Gamification Workshop Proceedings* (pp. 12–15). Vancouver, BC, Canada. Retrieved from http://www.mendeley.com/research/gamification-toward-definition/

Díaz, P., Aedo, I., & Rosson, M. B. (2008). Visual representation of web design patterns for end-users. *Proceedings of the Working Conference on Advanced Visual Interfaces - AVI '08*, 408. doi:10.1145/1385569.1385645

Donath, J., Karahalios, K., & Viegas, F. (1999). Visualizing conversation. *Journal of Computer- …*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.1999.tb00107.x/full

Dong, J., Yang, S., Zhang, K., & Member, S. (2007). Visualizing Design Patterns in Their Applications and Compositions, *33*(7), 433–453.

Dror, G., Koren, Y., Maarek, Y., & Szpektor, I. (2011). I Want to Answer , Who Has a Question ? Yahoo ! Answers Recommender System. *Work*, 1109–1117.

Duffy, T. M., Dueber, B., & Hawley, C. L. (1998). Critical thinking in distributed environment: A pedagogical base for the design of conferencing systems. In *Electronic Collaborators: Learner-centered technologies for literacy, apprenticeship and discourse* (pp. 51–78).

Durdik, Z., & Reussner, R. (2013). On the appropriate rationale for using design patterns and pattern documentation. *Proceedings of the 9th International ACM Sigsoft Conference on Quality of Software Architectures - QoSA '13*, (1), 107. doi:10.1145/2465478.2465491

Dutson, A. J., Todd, R. H., Magleby, S. P., & Sorensen, C. D. (1997). A review of literature on teaching engineering design through project-oriented capstone courses. *Journal of Engineering Education*, *86*(January), 17–28. Retrieved from http://jee.org/1997/january/503.pdf

Eades, P. (1984). A Heuristic for Graph Drawing. *Congressus Numerantium*, *42*(11), 149–160.

Engdahl, B., Köksal, M., & Marsden, G. (2005). Using treemaps to visualize threaded discussion forums on PDAs. *CHI'05 Extended Abstracts on Human …*, 1–4. Retrieved from http://dl.acm.org/citation.cfm?id=1056915

Farzan, R., DiMicco, J., & Millen, D. (2008). Results from deploying a participation incentive mechanism within the enterprise. *Proceedings of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, 563–572. Retrieved from http://dl.acm.org/citation.cfm?id=1357145

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, *7*(2), 117–140. doi:10.1177/001872675400700202

Fichman, P. (2011). A comparative assessment of answer quality on four question answering sites. *Journal of Information Science*, *37*(5), 476–486. doi:10.1177/0165551511415584

Fortuna, B., Grobelnik, M., & Mladenić, D. (2005). Visualization of Text Document Corpus, *29*, 497–502.

Franchetti, M., & Hefzy, M. (2012). Framework for Implementing Engineering Senior Design Capstone Courses and Design Clinics. *Journal of STEM Education …*, *13*(3), 30–45. Retrieved from http://eric.ed.gov/?id=EJ988059

Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). *Design Patterns: Elements of Reusable Object-Oriented Software*.

Gibbs, W. J., Olexa, V., & Bernas, R. S. (2006). A Visualization Tool for Managing and Studying Online Communications, *9*, 232–243.

Giguet, E., & Lucas, N. (2009). Creating discussion threads graphs with Anagora. In *Proceedings of the 9th international conference on COMPUTER SUPPORTED COLLABORATIVE LEARNING PRACTICES* (pp. 616–620). Retrieved from http://dl.acm.org/citation.cfm?id=1600141

Ginat, D. (2009). Interleaved pattern composition and scaffolded learning. *ACM SIGCSE Bulletin*, *41*(3), 109. doi:10.1145/1595496.1562915

Girgensohn, A., & Lee, A. (2002). Making web sites be places for social interaction. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work* (pp. 136–145). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=587098

Google Answers. (n.d.). Retrieved from http://answers.google.com/answers/

Harper, F. M., Raban, D., Rafaeli, S., & Konstan, J. A. (2008). Predictors of answer quality in online Q&A sites. *Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems - CHI '08*, 865. doi:10.1145/1357054.1357191

Heer, J., Card, S. K., & Landay, J. A. (2005). Prefuse: a toolkit for interactive information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (p. 421). New York, New York, USA: ACM Press. doi:10.1145/1054972.1055031

Heer, J., Shneiderman, B., & Park, C. (2012). A taxonomy of tools that support the fluent and flexible use of visualizations. In *Interactive Dynamics for Visual Analysis* (Vol. 10, pp. 1–26). doi:10.1145/2133416.2146416

Hsieh, G., & Counts, S. (2009). mimir : A Market-Based Real-Time Question and Answer Service. *Human-Computer Interaction*, 769–778.

Hsieh, G., Kraut, R. E., & Hudson, S. E. (2010). Why pay?: exploring how financial incentives are used for question & answer. In *Proceedings of the 28th international conference on Human factors in computing systems* (pp. 305–314). New York, New York, USA: ACM. doi:10.1145/1753326.1753373

Huang, J., Dasgupta, A., Ghosh, A., Manning, J., & Sanders, M. (2014). Superposter behavior in MOOC forums. *Proceedings of the First ACM Conference on Learning @ Scale Conference - L@S '14*, 117–126. doi:10.1145/2556325.2566249

Hung, W., Jonassen, D. H., & Liu, R. (2012). *Encyclopedia of the Sciences of Learning*. (N. M. Seel, Ed.)*Encyclopedia of the Sciences of Learning*. Boston, MA: Springer US. doi:10.1007/978-1-4419-1428-6

Indratmo, J. V., & Gutwin, C. (2008). Exploring blog archives with interactive visualization. In *AVI 2008* (pp. 39–46). doi:10.1145/1385569.1385578

Jain, S., Chen, Y., & Parkes, D. C. (2009). Designing incentives for online question and answer forums. In *Proceedings of the 10th ACM conference on Electronic commerce* (pp. 129–138). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1566374.1566393

Jyothi, S., McAvinia, C., & Keating, J. (2012). A visualisation tool to aid exploration of students' interactions in asynchronous online communication. *Computers & Education*, *58*(1), 30–42. doi:10.1016/j.compedu.2011.08.026

Karau, S., & Williams, K. (1993). Social Loafing: A Meta-Analytic Review and Theoretical Integration. *Journal of Personality and Social Psychology*, *65*(4), 681–706. doi:citeulike-article-id:1260763

Kim, B., Johnson, P., & Baker, J. (2006). Empirical evaluation of a visual interface for exploring message boards. In *Proceedings of the Second international conference on Advances in Visual Computing - Volume Part I* (pp. 293–302).

Kim, S., & Kim, S. (2012). A Model for Popularity Dynamics to Predict Hot Articles in Discussion Blog.

Köppe, C., & Utrecht, H. (2011). A pattern language for teaching design patterns. *Proceedings of the 18th Conference on Pattern Languages of Programs - PLoP '11*, (Part 2), 1–16. doi:10.1145/2578903.2579161

Lam, F., & Donath, J. (2005). Seascape and Volcano : Visualizing Online Discussions Using Timeless Motion, 1585–1588.

Lampe, C., Wash, R., Velasquez, A., & Ozkaya, E. (2010). Motivations to participate in online communities. *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10*, 1927. doi:10.1145/1753326.1753616

Lee, C. B., Garcia, S., & Porter, L. (2013). Can peer instruction be effective in upper-division computer science courses? *ACM Transactions on Computing Education*, *13*(3), 1–22. doi:10.1145/2499947.2499949

Lee, Y. (2012). IssueGalaxy : Interactive Visualizing Tools for the Current Issues in an Internet Discussion Site, 1–5. doi:10.1145/2184751.2184800

Leuski, a., & Allan, J. (2000). Lighthouse: showing the way to relevant information. *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*, *2000*, 125–129. doi:10.1109/INFVIS.2000.885099

Lin, J., & Landay, J. A. (2008). Employing Patterns and Layers for Early-Stage Design and Prototyping of Cross-Device User Interfaces, 1313–1322.

Mamykina, L., Manoim, B., Mittal, M., Hripcsak, G., & Hartmann, B. (2011). Design lessons from the fastest q&a site in the west. In *Proceedings of the 2011 annual conference on Human factors in computing systems* (pp. 2857–2866). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1979366

Massie, D. D., & Massie, C. A. (2006). Framework for organization and control of capstone design/build projects. *Journal of STEM Education*, *7*(3), 36–43.

Mazza, R., & Dimitrova, V. (2005). Generation of graphical representations of student tracking data in course management systems. In *IProceedings of the Ninth International Conference on Information Visualisation (IV'05)*. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1509087

Mazza, R., & Milani, C. (2005). Exploring usage analysis in learning systems: gaining insights from visualisations. In *12th International Conference on Artificial Intelligence in Education (AIED 2005)* (pp. 65–72).

McCallum, A. K. (2002). MALLET: A Machine Learning for Language Toolkit. Retrieved from http://mallet.cs.umass.edu/

Merriam-Webster Dictionary. (n.d.). Retrieved from http://www.merriam-webster.com

Muller, O. (2005). Pattern oriented instruction and the enhancement of analogical reasoning. *Proceedings of the 2005 International Workshop on Computing Education Research - ICER '05*, 57–67. doi:10.1145/1089786.1089792

Muller, O., Ginat, D., & Haberman, B. (2007). Pattern-oriented instruction and its influence on problem decomposition and solution construction. *ACM SIGCSE Bulletin*, *39*(3), 151. doi:10.1145/1269900.1268830

Nam, K. K., Arbor, A., & Ackerman, M. S. (2009). Questions in , Knowledge iN ? A Study of Naver ' s Question Answering Community University of Michigan. *Human Factors*, 779–788.

Pera, M. S. (2011). A Community Question-Answering Refinement System. *Strategy*, 251–260.

Perry, E., & Donath, J. (2004). Anthropomorphic Visualization : A New Approach for Depicting Participants in Online Spaces, 1115–1118.

Petter, S., Khazanchi, D., & Murphy, J. D. (2010). A design science based evaluation framework for patterns. *ACM SIGMIS Database*, *41*(3), 9. doi:10.1145/1851175.1851177

Piazza. (n.d.). Retrieved January 1, 2015, from http://www.piazza.com

Pomerantz, J., Nicholson, S., Belanger, Y., & David Lankes, R. (2004). The current state of digital reference: validation of a general digital reference model through a survey of digital reference services. *Information Processing & Management*, *40*(2), 347–363. doi:10.1016/S0306-4573(02)00085-7

Raymer, R. (2011). Gamification: Using Game Mechanics to Enhance eLearning. *eLearn*. doi:10.1145/2025356.2031772

Richardson, M., & White, R. W. (2011). Supporting synchronous social Q&A throughout the question lifecycle. In *Proceedings of the 20th international conference on World wide web* (pp. 755–764). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1963511

Roberts, L., & Roberts, T. L. (1998). Are newsgroups virtual communities? In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 360–367). ACM Press/Addison-Wesley Publishing Co. Retrieved from http://dl.acm.org/citation.cfm?id=274694

Saraiya, P., Shaffer, C. a., McCrickard, D. S., & North, C. (2004). Effective features of algorithm visualizations. *ACM SIGCSE Bulletin*, *36*(1), 382. doi:10.1145/1028174.971432

Schweitzer, D., & Brown, W. (2007). Interactive visualization for the active learning classroom. *ACM SIGCSE Bulletin*, *39*(1), 208. doi:10.1145/1227504.1227384

Seaton, D. T., Bergner, Y., Chuang, I., Mitros, P., & Pritchard, D. E. (2014, April 1). Who does what in a massive open online course? *Communications of the ACM*, *57*(4), 58–65. doi:10.1145/2500876

Sebrechts, M., & Cugini, J. (1999). Visualization of search results: a comparative evaluation of text, 2D, and 3D interfaces. *Proceedings of the …*, 3–10. Retrieved from http://dl.acm.org/citation.cfm?id=312634

Sheard, J. (2004). Electronic learning communities: Strategies for establishment and management. *SIGCSE Bulletin (Association for Computing Machinery, Special Interest Group on Computer Science Education)*, *36*(3), 37–41. doi:10.1145/1026487.1008009

Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages*. doi:10.1109/VL.1996.545307

Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science (New York, N.Y.)*, *323*, 122–124. doi:10.1126/science.1165919

Sonnino, E. (2013). 4 Ways to Get the Most out of a MOOC. *Learning Advisor*. Retrieved September 10, 2014, from http://blog.studentadvisor.com/four-ways-get-mooc/

Spence, I., & Kutlesa, N. (1999). Using Color to Code Quantity in Spatial Displays. *Journal of Experimental Psychology*, *5*(4), 393–412.

StackOverflow. (n.d.). StackOverflow. Retrieved from http://www.stackoverflow.com

Suryanto, M. A., Lim, E. P., Sun, A., & Chiang, R. H. L. (2009). Quality-aware collaborative question answering. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining - WSDM '09* (pp. 142–151). New York, New York, USA: ACM Press. doi:10.1145/1498759.1498820

Szekely, P. (1996). Retrospective and challenges for model-based interface development. In *3rd Int. Workshop on Computer-Aided Design of User Interfaces*. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Retrospective+and+Challenges+for+Model-Based+Interface+Development#0

Tidwell, J. (2005). *Designing Interfaces: Patterns for Effective Interaction Design*. *OReilly Media Inc* (Second.). O'Reilly Media. Retrieved from http://www.amazon.com/Designing-Interfaces-Patterns-Effective-Interaction/dp/0596008031

Trampuš, M., & Grobelnik, M. (2010). Visualization of online discussion forums. *Workshop on Pattern Analysis Applications*, *11*, 134–141. Retrieved from http://jmlr.org/proceedings/papers/v11/trampus10a/trampus10a.pdf

Vanderdonckt, J. (2010). Generative Pattern-Based Design of User Interfaces, 12–19.

Wang, G., Gill, K., & Mohanlal, M. (2013). Wisdom in the social crowd: an analysis of quora. *Proceedings of the 22nd …*, 1341–1351. Retrieved from http://tech.sina.com.cn/i/2014-05-24/22489398312.shtml\nhttp://dl.acm.org/citation.cfm?id=2488506

Wanner, F., Ramm, T., & Keim, D. (2011). ForAVis: explorative user forum analysis. In *WIMS '11: Proceedings of the International Conference on Web Intelligence, Mining and Semantics*. Retrieved from http://dl.acm.org/citation.cfm?id=1988705

Warren, J., Rixner, S., Greiner, J., & Wong, S. (2014). Facilitating Human Interaction in an Online Programming Course. *Proceedings of the 45th ACM Technical Symposium on Computer Science Education - SIGCSE '14*, 665–670. doi:10.1145/2538862.2538893

Wattenberg, M., & Millen, D. (2003). Conversation thumbnails for large-scale discussions. *CHI '03 Extended Abstracts on Human Factors in Computer Systems - CHI '03*, 742. doi:10.1145/765962.765963

Wilkinson, B., Villalobos, J., & Ferner, C. (2013). Pattern programming approach for teaching parallel and distributed computing. *Proceeding of the 44th ACM Technical Symposium on Computer Science Education - SIGCSE '13*, 409. doi:10.1145/2445196.2445319

Xia, H., Li, Z., Wang, H., & Gu, Y. (2007). A Lightweight Method of Web Service Ontology Merging Based on Concept Lattice. *The 2nd IEEE Asia-Pacific Service Computing Conference (APSCC 2007)*, 304–311. doi:10.1109/APSCC.2007.49

Xiong, R., & Donath, J. (1999). PeopleGarden: creating data portraits for users. *Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology*, 37–44. doi:10.1145/320719.322581

Yahoo! Answers. (n.d.). Retrieved from http://answers.yahoo.com/

Zegarra, E., Chang, S.-K., & Wang, J. (2015). Motivating Participation in Educational Settings via an Improved Q & A System. In *2015 International Workshop on Distance Education Technologies (accepted)*.

Zegarra, E., Colace, F., De Santo, M., & Chang, S.-K. (2010). An Ontology-based Configurator for Customized Product Information based upon the Slow Intelligence Systems Approach. In *Proceedings of the 22nd International Conference on Software Engineering Knowledge Engineering (SEKE'2010)* (pp. 521–528).