Department of Informatics
University of Fribourg (Switzerland)

THE FORA FRAMEWORK
A FUZZY GRASSROOTS ONTOLOGY FOR ONLINE REPUTATION MANAGEMENT

PhD Thesis

Submitted to the Faculty of Sciences at the University of Fribourg (Switzerland) to obtain the degree of Doctor scientiarium informaticarum

by

Eduard Karl PORTMANN

from Ruswil LU (Switzerland)

Thesis N° 1742
UniPrint
2012

Accepted by the Faculty of Sciences of the University of Fribourg (Switzerland), on the recommendation of:

- *Professor Ulrich Ultes-Nitsche*: University of Fribourg (Switzerland) as Jury President
- *Professor Andreas Meier*: University of Fribourg (Switzerland) as PhD Director
- *Professor Philippe Cudré-Mauroux*: University of Fribourg (Switzerland) as Internal Expert
- *Professor Witold Pedrycz*: University of Alberta, Edmonton (Canada) as External Expert

Fribourg, February 6, 2012

PhD Director

Professor Andreas Meier

Faculty Dean

Professor Rolf Ingold

To my love Eveline and our little princess Shani.
The whole in the part.

ACKNOWLEDGEMENTS

This thesis grew out of years of argument. I thank my many friends, colleagues, diploma and fellow doctoral students, advisors, and my mentor for their time and ideas and for the chance to debate them.

First of all, I would like to thank my mentor, first advisor, and friend Andreas Meier. You gave me the idea and the opportunity of writing this thesis. Andreas, for me you are one of the most competent people I have met in my life, both personal as well as professional. I am deeply grateful and proud to have been a student of yours. I thank you very much for this!

I also would like to thank my thesis co-advisors Philippe Cudré-Mauroux and Witold Pedrycz for agreeing to be members of my PhD committee, as well as Ulrich Ultes-Nitsche for his commitment as president of it. You supported me always with professional assistance, careful reading and valuable feedback. Besides I also owe thank to my advisors Lotfi Zadeh for pointing me in the right directions, Adrian Cheok for giving me the chance to collaborate, and Diana Ingenhoff for introducing me into reputation and issue management. Thanks a million.

Warm thanks go to my fellow doctoral students, Dani Fasel, Darius Zumstein, Joël Vogt, Luis Terán, Marco Savini, Michael Kaufmann, and Zoltán Horváth. Writing a PhD thesis is a long and lonely way that I would not have faced without your presence. You are a great crowd.

I am also thankful to my former diploma students, Andrea Liechti, Aron Martinez, David Oggier, Fernando Belfort, Hanspeter Siegfried, Katrin Uhlmann, Livia Hächler, Marc Osswald, Nicolas Spycher, Pascal Burkhard, Roger Fuchs, Sandro Kolly, and Simon Marti. Your Licentiate, Master, or Bachelor thesis made a significant contribution to this PhD thesis. I like all of your work.

Special thanks goes to Anthon, Anita, Jigé, Joël, Luis, Lukas, Marcel, Marcin, Marco, Mike, Philipp, Roland, Sylviane, Tam, and Urs, for keeping things running and always being helpful.

Last I would like to thank all of my numerous friends who can be considered rightful coauthors. Thank you all for discussing my ideas.

Most of this thesis was written during a peripatetic period when I freed myself of (almost) all obligations, routines, and pressures. I wrote it largely in my contemporary office in Fribourg but I also spent much time at home absorbed in my writing to the point that I forgot taking care of my love Eveline and our little princess Shani. I express regret for that and thank you both for your patience and support.

Edy Portmann
February 2012

ABSTRACT

Online reputation management deals with monitoring and influencing the online record of a person, an organization or a product. The Social Web offers increasingly simple ways to publish and disseminate personal or opinionated information, which can rapidly have a disastrous influence on the online reputation of some of the entities.

This dissertation can be split into three parts: In the first part, possible fuzzy clustering applications for the Social Semantic Web are investigated. The second part explores promising Social Semantic Web elements for organizational applications, while in the third part the former two parts are brought together and a fuzzy online reputation analysis framework is introduced and evaluated.

The entire PhD thesis is based on literature reviews as well as on argumentative-deductive analyses. The possible applications of Social Semantic Web elements within organizations have been researched using a scenario and an additional case study together with two ancillary case studies—based on qualitative interviews. For the conception and implementation of the online reputation analysis application, a conceptual framework was developed. Employing test installations and prototyping, the essential parts of the framework have been implemented.

By following a design sciences research approach, this PhD has created two artifacts: a framework and a prototype as proof of concept. Both artifacts hinge on two core elements: a (cluster analysis-based) translation of tags used in the Social Web to a computer-understandable fuzzy grassroots ontology for the Semantic Web, and a (Topic Maps-based) knowledge representation system, which facilitates a natural interaction with the fuzzy grassroots ontology. This is beneficial to the identification of unknown but essential Web data that could not be realized through conventional online reputation analysis.

The inherent structure of natural language supports humans not only in communication but also in the perception of the world. Fuzziness is a promising tool for transforming those human perceptions into computer artifacts. Through fuzzy grassroots ontologies, the Social Semantic Web becomes more naturally and thus can streamline online reputation management.

KURZFASSUNG

Das Online Reputationsmanagement beschäftigt sich mit der Überwachung und Beeinflussung des Online Rufs einer Person, einer Organisation oder eines Produkts. Da das Soziale Web einfache Möglichkeiten bietet, reputationsrelevante Einträge zu veröffentlichen und diese häufig rasante Weiterverbreitung finden, können negative Einträge rufschädigende Auswirkungen haben.

Diese Dissertation kann in drei Teile aufgeteilt werden: Im ersten Teil werden Anwendungsmöglichkeiten der fuzzy Clusteranalyse für das Soziale Semantische Web untersucht, im zweiten Teil Einsatzmöglichkeiten von Sozialen Semantischen Web Elementen in Organisationen ausgelotet und im dritten Teil werden die vorherigen zwei Teile zusammengeführt und ein Framework und dessen Implementation für eine fuzzy Online Reputationsanalyse vorgestellt und evaluiert.

Der gesamten Dissertation liegen Literaturanalysen und argumentative-deduktive Analysen zugrunde. Die Einsatzmöglichkeiten von Sozialen Semantischen Web Elementen in Organisationen wurden mittels eines Szenarios und einer darauf aufbauenden Fallstudie, sowie zwei ergänzenden (auf qualitativen Interviews beruhenden) Fallstudien erforscht. Für die Herleitung und Implementation der Online Reputationsanalyse Applikation wurde ein konzeptionelles Framework entwickelt. Mittels Testinstallationen und Prototyping wurden die wesentlichen Teile des Frameworks umgesetzt.

Einem Design Science Forschungsansatz folgend wurden in diesem Dissertationsprojekt zwei Artefakte erstellt: Das Framework sowie ein Prototyp als Machbarkeitsnachweis. Beiden Artefakten liegen zwei Kernelemente zugrunde: Eine (clusteranalysebasierte) Übersetzung von im Sozialen Web benutzten Schlagwörter zu einer computerverständlichen fuzzy Basisontologie für das Semantische Web und ein (Topic-Maps-basiertes) Wissensrepräsentationssystem, welches eine natürliche Interaktion mit der fuzzy Basisontologie erlaubt. Dies fördert die Entdeckung unbekannter aber wesentlicher Webquellen, welche mittels herkömmlicher Online Reputationsanalyse nicht gefunden werden könnten.

Die natürlicher Sprache inhärente Struktur dient Menschen nicht nur als Kommunikationsmittel, sondern auch als Wahrnehmungshilfe. Fuzziness stellt eine erfolgversprechende Methode der Übermittlung dieser menschlichen Wahrnehmungen an Computer dar. Das Soziale Semantische Web wird mittels der fuzzy Basisontologie natürlicher und kann so das Online Reputationsmanagement optimieren.

RÉSUMÉ

La gestion de la réputation en ligne est responsable pour la surveillance et pour l'influence des enregistrements en ligne d'une personne, d'une organisation ou d'un produit. Les réseaux sociaux permettent désormais facilement de publier très rapidement des informations ayant un impact sur la réputation; si ces données sont négatives, elles peuvent dramatiquement nuire à certaines de ces entités.

Ce travail de thèse se divise en trois parties: la première partie examine les possibilités d'application du partitionnement flou dans le domaine du Web sémantique et social, la deuxième partie explore des éléments Web sémantique et social promettant pour les applications organisationnelles et la troisième partie réunit les deux premières en présentant un ensemble d'outils conceptuels ainsi qu'un prototype d'application de gestion de la réputation en ligne et une évaluation de ce dernier.

Cette thèse est basée sur des évaluations critiques de la littérature et sur des analyses d'argumentation et de déduction. Les possibilités d'applications du Web sémantique et social dans des organisations ont été analysées par le biais d'un scénario et de trois études de cas dont deux basées sur des interviews qualitatives. Un ensemble d'outils conceptuels a été développé pour la conception et l'implémentation de l'application d'analyse de réputation en ligne. Les parties essentielles de ces outils ont été mises en œuvre via l'implémentation d'un prototype.

Suivant une approche scientifique, deux produits ont été conçus comme preuve de la viabilité de thèse: un ensemble d'outils conceptuels et un prototype d'application d'analyse de réputation en ligne. Ces produits se basent sur deux éléments. Le premier est la traduction (basée sur des partitionnements flous) des mots-clés utilisés dans le Web sémantique et social via une ontologie floue compréhensible par un système d'information. Le deuxième est un système de représentation du savoir basé sur des cartes topiques qui facilitent une interaction naturelle avec l'ontologie. Cela favorise la découverte de sources sur le Web, inconnues mais essentielles, qui ne peuvent pas être trouvées par le biais d'analyses de réputations en ligne traditionnelles.

La structure inhérente du langage naturel de l'homme n'est pas qu'un moyen de communication mais aussi un moyen de perception. Le flou applique cette perception humaine à l'ordinateur. En utilisant une ontologie floue, le Web sémantique et social devient plus naturel et peut ainsi optimiser la gestion de la réputation en ligne.

CONTENTS

LIST OF ALGORITHMS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACID	Atomicity, Consistency, Isolation, and Durability
ANT	Actor Network Theory or Another Neat Tool
API	Application Programming Interface
APP	Atom Publishing Protocol
ARM	Approximate Reasoning Methods
BLD	Basic Logic Dialect
CCO	Chief Communication Officer
CEO	Chief Executive Officer
CMS	Content Management System
CRM	Customer Relationship Management
CSO	Cluster Supporting Objects
CSS	Cascading Style Sheets
CWO	Chief Web Officer
DL	Description Logic
DOAP	Description Of A Project
FCM	Fuzzy C-Means
FLAME	Fuzzy clustering by Local Approximation of MEmberships
FOAF	Friend-Of-A-Friend
FORA	Fuzzy Online Reputation Analysis
GK	Gustafson-Kessel
GUI	Graphical User Interface
HCI	Human-Computer Interaction
HCIR	Human-Computer Information Retrieval
HITS	Hyperlinked-Induced Topic Search
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
IR	Information Retrieval
InRiNa	Innovation- and Risk-Navigator
ISO	International Standardization Organization
JSON	JavaScript Object Notation
KAON	KArlsruhe ONtology
KNN	k-Nearest Neighbor
NLP	Natural Language Processing
NUS	National University of Singapore
OWA	Open-World Assumption
OWL	Web Ontology Language
PDA	Personal Digital Assistant
PHP	PHP Hypertext Preprocessor
PR	Public Relations
PRD	Production Rule Dialect
RAM	Random-Access Memory
RDF	Resource Description Framework
RDFa	RDF in Attributes
RDFS	RDF Schema

REST	REpresentational State Transfer
RIF	Rule Interchange Format
RSS	Really Simple Syndication
SEO	Search Engine Optimization
SERP	Search Engines Result Pages
SIOC	Semantically-Interlinked Online Communities
SME	Small and Medium Enterprises
SMS	Short Messages Service
SNSF	Swiss National Science Foundation
SPARQL	SPARQL Protocol and RDF Query Language
SQL	Structured Query Language
TMMN	Topic Maps Martian Notation
UGC	User-Generated Content
UML	Unified Modeling Language
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
WEF	World Economic Forum
WOM	Word Of Mouth
WOT	Web of Things
WSDL	Web Services Description Language
WWW	World Wide Web or Web for short
XML	eXtensible Markup Language
XTM	XML Topic Maps

LIST OF SYMBOLS

A	Fuzzy Set
β	Recall Parameter
C	Cluster Prototype
c_i	Centroid of Cluster i
c	Cluster Number
Γ	Cluster of Data Partitions
d	Distance
d'	Diameter
E	Local (Neighborhood) Approximation Error
FN	False Negative
FP	False Positive
F_β	F-Measure
g	Probabilistic Data Partitioning Function
I_D	Dunn Index
I_{DB}	Davies-Bouldin Index
I_{fDB}	fuzzy David-Bouldin Index
I_R	Rand Index
μ	Membership Function
J	Objective Function
K	Search Weight
L_p	L_p or Minkowski Distance
p	L_p Constant
m	Weightening Exponent
n	Number of Objects
P	Precision
Π	Proposition
q	Terms
R	Recall
ρ	Density of a Point
$\mathbb{R}$	Real Numbers
U	Partition Matrix
u	Cluster Assignement
σ_i	Distance of Objects in i to c_i
TN	True Negative
TP	True Positive
τ	Iteration Number
u	Individual Object
w_{ij}	Weighting of i to j
X	Dataset
x	Object
Ω	Universe of Discourse
$\wedge$	Logical AND
$\vee$	Logical OR
$\neg$	Logical NOT

1 PREFACE

"A traveler without observation
is a bird without wings."

—Moslih Eddin Saadi

In the last few years, the World Wide Web (WWW), or Web for short, has increasingly evolved to a social venue. Its users' social activity has moved beyond message boards to become a wider part of the Web. The Social Web—including (micro-) blogging platforms such as Blogger[1], Twitter[2], and WordPress[3], content sharing platforms like Flickr[4], Last.fm[5], and Delicious[6], social networking platforms as Facebook[7], Google+[8], and LinkedIn[9], as well as wikis such as Wikipedia[10], Wikitravel[11] and Wikiquote[12]—has seized the attention of millions of users as well as billions of dollars in investment. A social website permits users to interact and cooperate with each other in a social media dialogue as consumers of User-Generated Content (UGC) in a virtual community, in contrast to websites where users are restricted to solely content viewing [Bell, 2009; Breslin et al., 2009].

Semantic Web representation mechanism on the other hand fit perfectly to express people, objects, and the relationships that bind them together in such object-centered networks, by recording and representing the heterogeneous ties that interconnects each to the other. Using agreed-upon standards to describe people, content objects, and the connections that interconnect them, these networks can also interoperate by appealing to common semantics. Unfortunately, using Semantic Web technologies is not that easy as using standard Social Web elements which is the reason why especially Web developers benefit from these technologies to augment the ways in which they create, reuse, and link content [Allemang & Hendler, 2008; Hitzler et al., 2010].

Hence the Social Semantic Web aims to complement the formal Semantic Web vision by adding a pragmatic approach relying on description lan-

guages for semantic browsing by using approximate classification and semiotic knowledge representations. Such a system has a continuous process to provide crucial domain knowledge through semi-formal taxonomies, folksonomies or ontologies. With the Social Semantic Web, the opportunity of human created loose semantics as a means to fulfill the vision of the Semantic Web is emphasized. Instead of relying completely on automated semantics with formal ontology handling and inferencing, humans collaboratively build semantics aided by information systems. While the Semantic Web enables information processing with automatic inferencing across domains, the Social Semantic Web opens up its doors for information systems on the Web, to add everyday semantics, thus allowing interoperability between objects, actions and their users [Blumauer & Pellegrini, 2009; Breslin et al., 2009; Zacklad et al., 2003].

The Social Semantic Web is a great research tool for hearing what is on consumers mind. With the explosion of social media, no organization, brand or individual escapes online mention by stakeholders. Hence, organizations can learn from the dialogue. It is a lot cheaper than, for example, holding focus groups in multiple cities. In contrast to focus groups, where consumers are interviewed, in the Social Web the consumers are the information providers themselves. More importantly, on these grounds the provided information is less biased by organization. Research shows that two-thirds of consumers never voice complaints directly to organizations [Beal & Strauss, 2008]. Yet, the Semantic Web technologies allow organizations to more easily identify these complaining discussions and through adept participation preserve a good reputation.

The following introductory chapter first provides to the reader in section 1.1 a motivation of the pertinence of the chosen pragmatic approach for the Social Semantic Web. Next, section 1.2 enumerates the research objectives that are treated in this thesis. Section 1.3 introduces the thesis underlying research methods of information system as well as computer and social science research. In section 1.4 an outline of the thesis is given. Finally, in section 1.5 some general information and published contributions which are part of this work are presented.

1.1 MOTIVATION OF RESEARCH

As already pointed out, in the last decade, so-called socio-semantic information systems have reformed the way information on the Web can be stored and managed. As a result, the information size has strongly accumulated thus leading frequently to information overload. It therefore becomes difficult to analyze the entire magnitude of available information and to generate appropriate management decisions.

In this context, this PhD thesis recommends the creation of fuzzy grassroots ontologies, which combines the Social Web with the Semantic Web. The creation of this kind of ontology is based on fuzzy clustering, which, in turn, is based on fuzzy logic and fuzzy set theory. Fuzzy set theory is an addition to traditional set theory and handles the concept of partial truth along with true and false, which is used for qualitative rather than quantitative judgment. Fuzziness follows the way humans think and helps to handle real world complexities more efficiently. Hence, it is useful in converting imprecise human information to precise mathematical models. It is shown that with fuzzy clustering it is possible to overcome the gap between the bottom-up-approach of Social Web's folksonomies and the top-down-approach of Semantic Web's ontologies. Since underlying fuzzy set theory is more suitable for handling vague information, it captures human vagueness and expresses it with adequate mathematical precision for computers to understand.

In this sense the Social Semantic Web can be seen as a Web of collective knowledge systems, which are able to provide useful information based on human contributions and which get better as more people participate. Thus the created fuzzy grassroots ontologies represents knowledge for computers as well as for humans and in doing so make possible inferencing (i.e. drawing conclusions) about a domain. Since knowledge is used to achieve intelligent behavior, the fundamental goal of knowledge representation is to present it in a manner which will facilitate reasoning — knowledge representation and reasoning being seen as two sides of the same coin.

Nevertheless, problem solving can be simplified by an appropriate choice of knowledge representation. Presenting it in the right ways makes certain problems easier to solve. Knowledge representation and reasoning may be used during a Web search process for example. Based on the fuzzy grassroots ontology, a Web search engine can find additional or better matching results or at least empower a user to interact with the Web search engine in a straightforward manner.

A pertinent and promising application field of the proposed fuzzy grassroots ontology is online reputation analysis. The Fuzzy Online Reputation Analysis or FORA (plural of forum, the Latin word for marketplace) framework is an abstraction for searching the Social Web to find meaningful information on reputation. The framework is based on a fuzzy grassroots ontology to carry

out online reputation management. Thereby fuzziness helps transforming vague human-provided information that appears in grassroots movements of the Social Web, to a computer-understandable ontology. In turn, among others, this computer-produced ontology can be used to manage online reputation with the purpose to monitor, address, and rectify undesirable mentions in the grassroots-impelled Social Web.

The framework was developed as a functional adoption of accumulated knowledge of Web and media skills during the realization of the PhD project. Based on the automatic, fuzzy-built ontology, this framework queries the social marketplaces of an organization for reputation, combines the retrieved results, and generates navigable Topic Maps. Using these interactive maps, the organizations communications operatives can afterwards zero in on precisely what they are looking for and discover unforeseen relationships between topics and tags. So using this framework it is possible to scan the Social Web for a name, product, brand, or combination thereof and determine query-related topic classes with related terms and thus identify hidden sources.

Building and maintaining profitable customer relationships are important issues in the field of electronic commerce since the Web these days enables a global market. The Social Web consists of social media elements that provide online prosumers (combination of producer and consumer) a free and easy means for interacting or collaborating with each other. Consequently, it is not surprising that the number of people who read weblogs (short: blogs), for example, at least once a month has grown rapidly in the past few years and is likely to increase further in the foreseeable future. Blogging gives people the ability to express their opinions and to start conversations about matters that affect their daily lives. These conversations strongly influence what people think about organizations and what products they purchase. The influence of these conversations on potential purchases is leading many organizations to strategically conduct blogosphere scanning. Through this scanning, it is possible to identify conversations that mention an organization, a brand, the name of high-profile executives, or particular products. Besides, this also allows detecting misused blog posts (e.g. by competitors) to harm an organization. Through participation in the conversations, the affected parties can improve the organization's image, mitigate damage to their reputation posed by unsatisfied consumers and critics, and cautiously promote their products.

1.2 Research Issues

Combining well-grounded academic research with practice-oriented applications is a common practice within information system, computer and social research: especially in the development of the Social Semantic Web it is highly anticipated. This PhD thesis consists of three parts of equal value: In the first part the theoretical foundations are elaborated, whereas in the se-

cond part this theoretical foundation comes into operation. In particular this second part employs the foundations to online reputation management. Based on the two previous parts, in the third part a unifying framework and corresponding prototype is elaborated. The prototype serves as a proof of concept. Hence, in this PhD project the following objectives are pursued:

1. The first, mostly theoretical objective is to review the evolution of the Web to a Social Semantic Web as well as the history of fuzzy logic, set theory and clustering. Furthermore, the scope of promising fuzzy applications to the Social (Semantic) Web is elicited (see chap. 2 and 3).

2. The second objective is to define how data from the Social Web can be structured using fuzzy clustering methods. This structure fuses humans and computers in terms of Human-Computer Information Retrieval (HCIR), since fuzzy sets are able to translate vague human concepts to computer-understandable models (see chap. 3).

3. The third objective is to define how the structured data (i.e. information) can be administered using the most promising Social Semantic Web knowledge administration system. The knowledge administration system is selected based on Web standards and recommendations (investigated by the first objective; see chap. 6).

4. The fourth objective is to define how (professional) media users (e.g. employees, communication operatives or online journalists) are using the Social (Semantic) Web and its search engines. Thereby it is learned what the media users expect from future Web and corresponding search engine (e.g. for online reputation management; see chap. 5).

5. The fifth objective consists of two parts with the purpose to characterize and specify the responsibility of online reputation management and in the course of this to develop a framework for online reputation analysis:

    I. The first part is to characterize and specify online reputation management as the domain of study to structure a unifying framework. The selected domain of study that structures this framework is defined by the problem actuality and the importance for (media) information systems. The results found by the literature review and qualitative interviews flow into this selected field as well (see chap. 4 and 5).

    II. In the second part of the objective a universal framework for online reputation analysis is developed and validated. So for the field of online reputation analysis, this framework is as specific and solution-oriented as possible (see chap. 6).

6. The sixth objectives also consist of two parts concerned with the evaluation, and with the development and validation of a free Web-based prototype:

    I. In this first part an appropriate (interactive) knowledge representation for structured (Social) Web information is evaluated. The user needs as well as the selected domain of study influenced this knowledge representation (see chap. 6).

    II. In the second part a free Web-based prototype is developed and validated for the selected application field as proof of concept. Thereby an emphasis is on the creation of a manageable and comfortable dashboard. The criteria to define manageable and comfortable arise from the user needs (previously investigated by the fourth objective; see chap. 7).

1.3 Research Methods

This PhD thesis is realized following a design science research methodology on real-world business problems. Therefore it aims first at creating innovative concepts (i.e. the FORA framework) which improve the human and organizational capabilities, and secondly, at evaluating these concepts by providing concrete instantiations (i.e. the YouReputation[13] prototype). It is based on the following information system, computer and social science research methods [Becker et al., 2009; Hevner & Chatterjee, 2010; Österle et al., 2010; Wilde & Hess, 2007]:

- *Literature review:* In a literature review the critical points of current knowledge including findings as well as theoretical and methodological contributions are reviewed. For the thesis literature reviews were always used as first step of knowledge acquisition (see chap. 2 to 7). Often they were completed by comparisons (see chap. 6 and 7).

- *Argumentative-deductive analysis:* This analysis was the basis for the PhD thesis. Therein the theoretical foundations are picked up and consequently theories, models, approaches and arguments are developed out of the contributions found by the literature review (see chap. 3 to 7).

- *Scenario:* A scenario is a narrative description of anticipated user and system interactions in the context of daily activity. In this PhD project, the Apple Inc.[14] scenario includes information about goals, expectations, motivations, actions and reactions of a fuzzy online reputation analysis system (see chap. 5).

- *Case studies:* They are based on an in-depth investigation of a single individual, group, or event; they are thereby either descriptive or explanatory. In the PhD project case studies are on one hand used to explore knowledge, which is rather in the sense of design research (see chap. 5).

On the other hand a case study is used to analyze the FORA framework and the YouReputation prototype in comparison with online reputation analysis frameworks and tools available on the market. Thereby the benefits and limitations of the framework and prototype are illustrated (see chap. 7).

- *Structured interviews:* In the thesis qualitative interviews are used to figure out how professional media workers (e.g. employees, communication operatives or online journalists) use Web search engines and online reputation analysis applications. These types of interviews are best suited for focus group studies in which it would be beneficial to compare responses in order to answer research questions (see chap. 5).

- *Test installations:* With test installations as a quality assurance for the PhD project, most common software concerning the FORA framework and the YouReputation prototype specifications, as for example knowledge administration systems, are installed and in doing so, compared, evaluated and tested (see chap. 6).

- *Conceptual framework:* Such a framework is used to present a preferred approach to an idea or thought. The FORA framework originates from a conceptual framework and epitomizes a reusable, skeletal, semi-complete modular framework that can be specialized to produce custom applications. It includes building blocks of services and components that are essential for constructing a feature-rich online reputation analysis application (see chap. 6).

- *Prototyping:* This method of software development leads rapidly to results and allows early feedback regarding of the suitability of a possible solution approach. The implemented YouReputation prototype is used as a basis proof of concept for the FORA framework (see chap. 7).

1.4 Structure of this Thesis

This thesis is organized in three main parts: First, the theoretical background; second, the field of application; and third, the framework and its implementation. At this point an introduction to the chapters is given and the logical chapter layout and flow is explained. The first part relates to the conceptual perspective of this work; in other words, it contains the theoretical background on which the rest of the thesis is built:

- *Chapter 2 – The Social Semantic Web*: This chapter presents the transition of the Web to a Social Semantic Web. Stages of this presentation are the Social Web (incl. a classification of its elements), the Semantic Web (incl. World Wide Web Consortiums (W3C) standardized components), and the Social Semantic Web (as an element where the previous presented elements flow together). This chapter answers the search for the evolution of the Web.

- *Chapter 3 – Fundamentals of Fuzzy Clustering Methods*: In this chapter, the concept of cluster analysis in general with an emphasis on the process of unsupervised learning tasks of clustering itself is exposed. Thereby proximity measurements, methods for defining the optimal number of cluster and also cluster validation methods are pronounced. Subsequently the general (hard) approach is enhanced by a more flexible (fuzzy) approach. Thereby the main concepts and mathematical notions of fuzzy logic, set theory and clustering are elaborated. In the end, this chapter answers the question how fuzzy logic can be combined with Social Semantic Web (see chap. 2). Thereby the concept of a fuzzy grassroots ontology will be presented as a solution to bridge the gap between the bottom-up-approach of Social Web's folksonomies and the top-down-approach of Semantic Web's ontologies.

The second part of this thesis investigates the field of application—the online reputation management. In 2011, the World Economic Forum (WEF)[15] defined online reputation management as a promising field with high future development opportunities. Among other criteria, this nomination was used as a determinative to immerse deeper into online reputation management.

- *Chapter 4 – Online Reputation Analysis*: In this chapter online reputation and its management are introduced as possible field of application of the fuzzy grassroots ontology. Online reputation management is the practice of managing the Web reputation of a person, brand or business, with the goal of suppressing negative mentions entirely, or pushing them lower on Search Engine Result Pages (SERP) to decrease their visibility. In this chapter the task of online reputation analysis is highlighted whereby the aspects of knowledge representation that are constantly growing in importance are introduced. The fundamental aim thereby is to visualize knowledge in a way that facilitates inferencing for both humans and computers. In the course of this it is shown how fuzzy grassroots ontology (see chap. 3) can be visualized as an appealing interactive Topic Map to help communication operatives zero in on precisely what they are looking for and discover unforeseen relationships between topics and tags on the Web. The scanned topics can then be monitored in the process afterwards.

- *Chapter 5 – Fuzzy Online Reputation Analysis Requirements:* This chapter presents first a scenario of anticipated features of a fuzzy online reputation analysis application. On this basis case studies are presented to approximate the output (e.g. frameworks or prototypes) of the design research. Through multiple case studies it is possible to accumulate supporting evidence, which can continue until theoretical saturation is reached. Because of that an in-depth analysis and two qualitative interview studies are presented within this chapter. The in-depth analysis investigates why it is useful for an organization to invest in social media elements. However, a first interview investigates the challenges of Web

search engines; a second investigates the challenges of online reputation management. In the end essential requirements for the FORA framework are drawn.

The third and final part of this thesis, the framework and its implementation, aims at proving the FORA framework and the applicability of the framework in a real world environment by presenting a concrete implementation. The framework is a type of intermediated theory that attempts to connect aspects of the previous proposed conceptual perspective. Furthermore, in this part, a summary concerning the application of the fuzzy grassroots ontology for online reputation analysis is drawn:

- *Chapter 6 – The Fuzzy Online Reputation Analysis Framework*: This chapter brings the chapters 4 and 5 together and thus draws up the FORA framework. It first starts with a conceptual framework for fuzzy online reputation analysis — a reputation management task conducted by communication operatives — and then emerges to a possible implementation. To this end, this chapter outlines the underlying conceptual framework. Then it describes its architecture and the component interactions. For an implementation several key components have to be checked against each other. To this end, comparisons of encouraging fuzzy clustering algorithms, as well as seminal knowledge administration and interactive knowledge representation systems are performed next. Then the chapter finally makes a mention of the implications for a promising implementation.

- *Chapter 7 – The YouReputation Prototype*: This chapter is the instantiation of the FORA framework. It reveals the YouReputation prototype (i.e. a blend of your and reputation), an implementation for fuzzy online reputation analysis. First the YouReputation kernel is explained, followed by the de facto implementations of the single modular-constructed parts of the YouReputation prototype. At that, a separation of concern between information acquisition and the use of information in systems is taken at hand. For the evaluation in design research it is promising to have an instantiation of the framework. Hence, the YouReputation prototype is intended as proof of concept for the FORA framework. To assess the prototype in turn, this chapter presents a cluster validation approach to test the fuzzy grassroots ontology and a comparison to benchmark the fuzzy online reputation analysis application to others.

- *Chapter 8 – Conclusion*: In this chapter the key aspects developed in this thesis are first summarized whereby a comparison with the research issues is performed. In addition, promising application domains as well as further evolutions are discussed. This phase is final of a specific research effort. It is the result of satisficing, that is, thought there are still deviations in the behavior of the artifact from (multiply) revised hypothetical

predictions, the results are judged to be good enough. Finally, the chapter ends with a daring outlook.

1.5 General Information

At the end of this PhD thesis a glossary is annexed to retrieve FORA framework-relevant terms. However, some basic knowledge concerning information systems, computer and social sciences is taken for granted and therefore the glossary is not exhaustive.

The Web-based YouReputation prototype developed during this PhD project is up at the Web address: `http://youreputation.org`. This prototype constitutes a simplified but operative implementation of the FORA framework as proof of concept.

Many aspects of the present project have been published in international conferences, journals and in handbooks [Andrushevich et al., 2011; Portmann, 2008; Portmann & Hutter, 2011; Portmann et al., 2012]. Contributions concerning the fuzzy clustering approach, its application to online reputation analysis and the implementation aspects have been published in [Portmann, 2009; Portmann & Meier, 2010; Portmann et al., 2010; Portmann, 2011a; Portmann et al., 2012; Wehrle & Portmann, 2012; Portmann & Kuhn, 2010].

Note as a last that for the sake of simplicity in the formulation of this thesis, the masculine formulation is chosen, but always address both sexes.

Part I

THEORETICAL BACKGROUND

2 THE SOCIAL SEMANTIC WEB

"You affect the world
by what you browse."

—Tim Berners-Lee

Nowadays the Web is omnipresent, reaching into almost everyone's life. More and more Web users do not switch off their devices all the time, continuously receiving and sending messages, frequently looking for information, now and then evaluating this information, and so on. The means to reach the Web do thereby not stop at personal computers, but increasingly also include mobile devices. More and more users are sharing information online, are working collaboratively on a topic, as well as maintaining their relationship in the Web [Alby, 2008]. All of this is so pervasive that it feels absolutely natural. Consequently it is not surprising that topics related to the Social Web are experiencing a surge of interest, both from the scientific community as well as the industry. However, apart from this and maybe also apart from the public perception, a complementary technological revolution takes place—the rising adaption of Semantic Web technologies. The Semantic Web is a vision that the present Web will eventually include the notion of meaning and become a metadata-rich Web where presently human-readable content will contain computer-understandable semantics [Berners-Lee et al., 2001].

Today information in the Web is raw material and currency at the same time. The possibility to extrapolate this information constitutes one of the core competencies of the future. The quest of the development ability and logistics of information is at the center of the discussion about utility and functionality of Semantic Web technologies. This discussion is controversial; skeptics of the Semantic Web dismiss the W3C standards and methods as too complex and technology-driven, as to have a chance in the grassroots-impelled Social Web. In return, the representatives of the Semantic Web community raise legitimate questions concerning alternative tools and

methods to get a grip of the information overload produced by precisely the bottom-up-processes of Social Web [Blumauer & Pellegrini, 2009]. Both fractions can exalt valid claims to their arguments. In the process, slowly the understanding grows that only combined forms of top-down with bottom-up-approaches represent a reasonable solution. Thereby this solution should include not only the technological feasibility but also social acceptance. The objective to exploit the Social Web, as well as the assignment of social media elements for collaborative enrichment of Web content with computer-understandable metadata, are impressive manifestations of a trend towards the Social Semantic Web. A denominating symptom of this development is the ongoing convergence between social media elements and Semantic Web technologies. Examples can be found in [Blumauer & Pellegrini, 2009] and [Breslin et al., 2009].

This chapter should come across as a theoretical introduction in preparation for the main focus of this PhD project, the FORA framework and its YouReputation prototype implementation. It is important to grasp underlying theory in order to gain awareness of online reputation analysis problems (see chap. 4 et seq.). Thus, this chapter is intended to shine a light on the Social Web as well as the Semantic Web. As introduction, section 2.1 briefly highlights the convergence of information and media sciences in the Web. Accordingly, in section 2.2 the most significant elements of Social Web are revealed. Section 2.3 introduces the vision of the Semantic Web and presents its basic technologies. In section 2.4 both presented parts are merged to engender the discussed Social Semantic Web. Section 2.5 closes this chapter with suggestions for further readings.

2.1 History on the Convergence of Information and Media

During the evolution of human civilization, new technologies allowed to keep evermore (semi) structured data (i.e. information) in diverse media forms. Hence, the invention of the Web, as well as the progression towards the Social Semantic Web can be pronounced by the need to get over the growing amount of information.

Indeed, the initial creation and recording of information took off with cave paintings some 32,000 years ago. One of the first expressions by letter was the Sumerian cuneiform script written on clay tablets. Caused by the papyrus rolls of the ancient Egyptians, the scrolls of Greeks and Romans, a very own dynamic evolved. The development of creation and distribution of text was even accelerated by the invention of the printing press [Portmann, 2008]. The invention of photography added another key form of media, followed up by the invention of the phonograph for sound recording and the capability to effectively create movies [Manovich, 2001]. This proliferation unleashed the sine qua non to collect and organize media objects. In earlier times, they were organized in libraries. These libraries were (and still are) centralized collections with categorical organization and indexing principles, whereby the knowledge organization of the stored information is mostly expert-based [Breslin et al., 2009].

About three millennia ago, the ancient Assyrians annotated clay tablets with small labels to make them easier to tell apart when they were filed in baskets or on shelves. The idea survived into the twentieth century in the form of the catalog cards that librarians used to record data (e.g. a book's title, author, subject, etc.) before library records were moved to computers [Gavirilis et al., 2008]. The actual books constituted the data; the catalogue cards comprised the metadata. The term meta comes from the Greek word that denotes alongside, with, after, next. Metadata can be thought of as data about data, and it commonly refers to descriptive structured data about (Web) resources that can be used to help support a wide range of operations. To minimize information overload and consequently allow faster information access, experts manually record metadata about books on catalog cards, for example. To assign the library analogy to a Web without metadata, every word in every page in every book must be indexed. Because such indexing will lag the growth and change in the Web, it often yields poor search results. With even some basic metadata, using the library analogy again there are books with categories, titles, descriptions, ratings, yielding a much better retrieval. Unfortunately the effort for this manual indexing is not satisfactory in the growing amount of information.

Vannevar Bush was one of the first to perceive that the dissemination of information and knowledge in diverse media forms had opened up new challenges that central archives—and the manual indexing mechanisms of traditional libraries—could not fulfill. Here, information is regarded as an

essential element to derive knowledge, but knowledge can additionally include facts and understanding gained through experience, education or reason. After the Second World War, Vannevar Bush suggested the Memex proto-hypertext system in which "*an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory.*" [Bush, 1945]. The core of his system declared a shift from a knowledge organization through experts to knowledge organization trough individuals. Later Doug Engelbart suggested the first hypertext system, what, in turn, led Ted Nelson to implement such a hypertext system with a simple Graphical User Interface (GUI). However, only Tim Berners-Lee was able to realize part of the vision: The hypertext system, alias the Web, which made global information access in the first place feasible.

The invention of the Web technically spurred new horizons for the public use of databases and networks. Innovative Web browsers subsequently allowed also laymen to access this resource. Increasingly, also organizations discovered the Web as a medium of communication between its diverse locations, departments, and stakeholders. It was therefore subsequently incorporated into part of the business.

Seen from the perspective of the producers and consumers, the Web bit by bit turned into an interactive medium. As already envisioned by [McLuhan & Nevitt, 1972; Toffler, 1980], the Web provided its prosumers, for the first time a free and easy means for interacting or collaborating with each other. [Jenkins, 2008] carried on McLuhan and Nevitts work and illustrated the cultural approximation of old and new media. His term convergence culture describes an emerging pattern of relations bringing together entertainment, advertising, brands, and consumers. So the burst of the dot-com bubble had no impact on the growing use of the Web. Rather improved business models and new offerings were developed. With the expansion of bandwidth and database-driven applications, it was possible to provide ever-greater amounts of information via the Web. This changed the perception of its nature to the extent that it was considered as a platform to store content.

The growing percentage of UGC challenged increasingly the exclusive state of established news portals and knowledge bases. Hence, the influence of this UGC in the pre-media space increased over time. The term pre-media space denotes the area in the verge of traditional media's news aggregation on their portals. In the context of anti-globalization movement's alternative public media projects were started. In particular, the rise of blogs and the emerging blogosphere fueled expectations for grassroots journalism. Blogs became popular as a new form of communication. Likewise the number of online communities and volunteer contributors accelerated drastically. Earlier blogs and communities were smiled at, but now they are accepted as open communication and documentation media. The same is true for social networks; end of September 2011, that is to say, the number of active Face-

book users is exceeding 800 million [Olivarez-Giles, 2011]. Wikipedia, as another example, occupied as a non-commercial project a crucial communication point. The blogosphere, in turn, revealed a new form of decentralized news propagation. Hence with the success of the blogosphere, social networks, and the Wikipedia project, the public, organizations, states, and established traditional (mass) media had to deal with innovative controlling models in the emerging Social Web. Apropos, at the time of writing, Wikipedia is suggested as world's first digital and global world cultural heritage site [Sooth & Schoneville, 2011].

2.2 Social Web Elements and their Classification

The Social Web contains social media that include communication and interactive tools. Communication tools typically handle the capturing, storing and presentation of communication, usually written but increasingly including audio and video as well. Interactive tools handle mediated interactions between a pair or group of users. They focus on establishing and maintaining a connection among users, facilitating the mechanics of conversation and talk. This section showcases the Social Web from a social sciences perspective. On this ground, a classification of social media elements (and applications) is offered. Since a basic understanding of these elements is essential to understand this PhD thesis, the different social media elements are shortly characterized next. This characterization list thereby is illustrative rather than exhaustive.

The Social Web portrays the Web as social media elements and applications, and describes how people socialize or interact with each other throughout the Web. Hence the Social Web can be described as people linked and networking with engaging Web content in a conversational and participatory manner. Therefore [Ebersbach et al., 2010] define that the Social Web is assembled of *"Web-based applications that support human information exchange, relationship building and its maintenance, communication and collaborative cooperation in a social or community context, and the data that emerge and the relationships between people who use these applications.*" This thesis builds on this definition for the Social Web as well. The focus thereby is on social and not on technical criteria. Additionally within this thesis, a distinction between element and application is made; the former is defined as a manifestation of social media, the later as instantiation of the social media elements in the form of a practical tool. Social media applications may well consist of technical parts, but these parts are not in the center within this thesis. However, many of these applications share social software characteristics like the ability to upload information, service-oriented design, open Application Programming Interfaces (API), and Web feeds—as the Atom Publishing Protocol (APP) and Really Simple Syndication (RSS). A deeper introduction to social software can be found in [Ebersbach et al., 2010].

The plethora of social media elements is almost endless. That is why on nearly any area of social life an appropriate community on the Web can be found. Accordingly it is useful to determine these elements for their purpose. To categorize them, three criteria are considered:

- *Collaboration*: This comes across as gathering and production of knowledge (e.g. information, statements, findings, ratings, etc.). Here, people are grouping around a topic to collaboratively edit it.
- *Information*: Where the focus is on the dissemination of information and knowledge. This can comprise of hypertext, links, uploaded files (e.g. text, pictures, images, videos, etc.), as well as just comments (e.g. opinions, insights, ratings, etc.).
- *Relationship*: Whereby the focus is on the building and caring of a relationship of interpersonal connections. This is about meeting other people virtually to obtain information or recover connections—likewise from the real world.

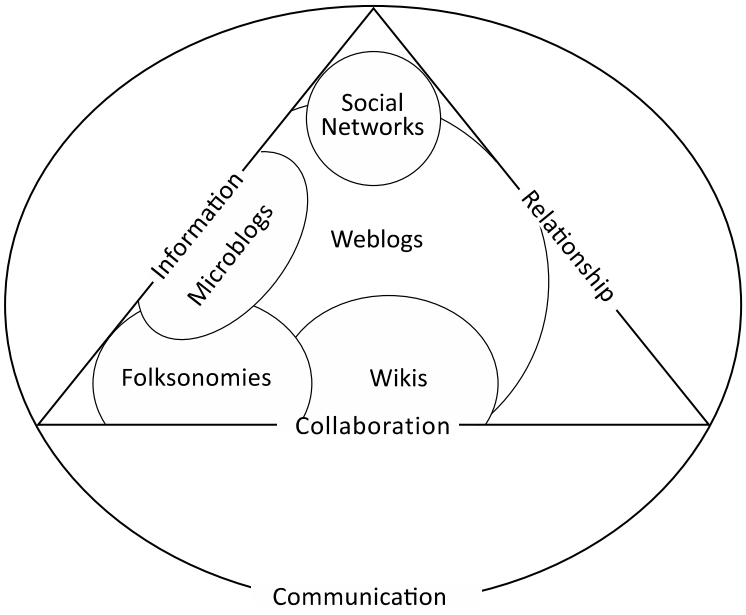

Figure 2.1: Triangular Classification Model.

Additionally, communication is an issue that is common to all elements of the Social Web in a more or less intense form. There are interactions between the different areas, but collaboration without communication, for example, is very difficult to imagine. The same applies to maintaining relationships and the exchange of information. So with communication the other three edges are kept together. Adapted from [Ebersbach et al., 2010], in figure 2.1 these edges (collaboration, information, and relationship) are visualized by a triangle. Since communication is concerned with all of these edges,

it is visualized as a circle around the triangle. Note that this figure is not intended as an exact mapping of social media elements but rather as an overview of them that flow into one another.

Nevertheless, an ideal medium that meets all three edges would be located in the middle of the triangle. In the following sections, the classified social media elements will be presented. Section 2.2.1 starts with weblogs. These are personal journals, typically administrated by individuals and updated on a daily basis with a personal view on a current issue. Microblogs differ from traditional blogs in that the content is typically smaller. They focus on short messages that are exchanged via a central platform. They have primarily a communicative character, and usually a short date range. They are presented in section 2.2.2. Section 2.2.3 introduces folksonomies (a blend of folk and taxonomy), which are systems of classification derived from the practice and method of collaboratively creating and managing tags to annotate and categorize content. Section 2.2.4 presents wikis, which focus on the collaborative creation of hypertexts, with the aim of the community to draw up content together. As a last element, in section 2.2.5, social networks are explained in more detail. They center on the development and maintenance of relationship. All of these elements increasingly can appear as a combination, such as wikis, with a social network extension—the talk is then about integrated platforms.

2.2.1 WEBLOGS

The expression weblog originates from the words Web and log and details a type of online diary (e.g. realized with WordPress or Blogger). According to [O'Reilly, 2005] blogging is a feature of the Social Web much sought-after. He observes blogging as one of the most common activities introduced by the Social Web and spot blogs as the mightiest media of UGC. Compared to websites, blogs are easier to handle and more flexible in their utilization. Through a blog Content Management System (CMS), information can be added and administrated straightforward. This simplicity is the basis for the prodigious and tremendously fast global expansion of weblogs.

The dialogue-based communication style, inherent to blogs, is an effect of the great number of links. Many blogs are interactive, allowing visitors to leave comments and it is this interactivity that distinguishes them from static or semi-static websites. In contrast to a static or semi-static website, an interactive website is one that changes frequently and automatically, based on certain criteria. Posts are connected and referenced by a trackback function. Trackbacks lead to other blogs that have been written about the same topic. Often they are reactions to the post they are linked to, but have been published in another blog instead of a comment. In order to work, trackbacks permalinks are necessary. This technology permits each entry to have its unique Web address through which it is retrievable at any time or place. By building bridges between blogs, permalinks turned them from an ease-of-

publishing phenomenon into a conversational medley of overlapping communities.

Blogs do not underlie a certain authority as the press, academia, medicine or law, which forms the interpretation of authenticity. The central characteristic of the blogosphere is that it flattens the different hierarchies; it equalizes the relationships through the fact that anyone may blog about anything [Hächler, 2010]. In the blogosphere traditional hierarchies have vanished and interactivity is central to it. Other ways of control seem to be emerging. Bloggers keep an eye on their audience; their main task is to manage the interaction and to keep it going. Analyzing received attention mainly does evaluating the importance of the different bloggers. Links, hits, trackbacks and being mentioned in other (top) blogs are indicators of value. Wanting to be spread and respected in the blogosphere makes bloggers careful when marking their own opinions.

[Levinson, 2009] stipulates two main characteristics of blogs: Firstly and as already introduced, anyone can blog about anything and secondly, the actual impact of a blog, as well as the time of its maximum impact, is incalculable. According to [Portmann, 2008], blogs can mobilize society to bring down politicians, hold an organization to account, popularize a book or spread a video, but blogs can also continuously echo a vicious lie, for example, long after it has been debunked [Myres, 2010].

Partaking in the blogosphere involves commenting. [Levinson, 2009] purports that comments are the most frequent form of sustained written discourse and attribute two other functions inherent to them: Firstly, comments can be an effective promotion for one's own blog. Secondly, and more important, comments do not solely go for the voice of the people, but also as the conveyors of the truth. They correct, if necessary, a post. Comments can therefore be a stumbling block for misuse of a blog (e.g. by competitors).

In business settings, weblogs are used either internal or external. The talk is then about corporate blogs. Such blogs are used by an organization to reach its structural goals. The benefit of corporate blogs is that posts and comments are easy to reach and follow. Corporate blogs are a connecting link between organizations and its customers.

In the subsequent section, a latest up-and-coming form of blogs will be presented: Microblogs attract additional bloggers afraid of text long entries.

2.2.2 MICROBLOGS

Microblogs (e.g. Twitter or Tumblr[16]) are a kind of revised form of traditional weblogs, where users can post short text messages, reporting on the details of one's life. The messages are typically restricted to 140-characters for compatibility with Short Messages Services (SMS) [Comm, 2009; O'Reilly & Milstein, 2009; Sagolla, 2009].

There exists also commercial microblogs to promote websites, services or products, and to push on collaboration within an organization [Portmann & Hutter, 2011]. Some microblogging services provide functionalities such as privacy settings, which permit to monitor who can read their microblogs, or other ways to promulgate post entries in addition to the Web-based interfaces, such as smartphones and Personal Digital Assistants (PDA).

Often microblogs are used to bring readers to the attention of traditional blog posts [Portmann & Hutter, 2011]. However, with social bookmarking platforms, these readers can also organize their bookmark-worthy Web content in a straightforward manner. Additionally these platforms can be used to annotate the Web content—folksonomies are such an example. Anyway, different microblogging systems even allow annotating its messages with hashtags. These hashtags are means to tag underlying message and may provide a folksonomy.

2.2.3 FOLKSONOMIES

Metadata can be used to provide a structured description of characteristics such as the meaning (i.e. semantics), content, structure and purpose of a Web resource; to facilitate information sharing; to enable more sophisticated search engines on the Web; to support intelligent agents and the pushing of data (e.g. from Web feeds); to minimize data loss or repetition; and to help with the discovery of resources by enabling field-based searches.

The interactive and participative possibilities of the Social Web also have their effect on the way people organize and share their online sources. Tagging and the emerging folksonomies are the result of people describing or labeling Web content (e.g. bookmarks on Delicious or images on Flickr). [Smith, 2008] characterizes folksonomy as the popular term describing the bottom-up classification systems that emerge from collaborative tagging content. Tags can be seen as metadata about a resource. They can be used in various situations. For example, an image platform allows one to upload photos, sort them, and subsequently organize them through tags. On microblogging platforms, as another example, often the possibility to annotate content by hashtags is provided (see sect. 2.2.2). Social bookmarking platforms, in turn, entail being able to add tags to a user's bookmarks (e.g. links, pictures, movies, etc.). By adding whatever keywords suit best, each user creates his collection of links that are being categorized through these keywords. Whenever a user finds a website that is meaningful to save or mark, he can do so by describing it through keywords or tags. This way he adds metadata to the online source, creating multiple and especially personal ways of finding it again. In addition, the tags and sources are being shared with the entire community, thus enabling to pinpoint new resources including the same or similar tags. Tagging improves the findability of resources by using individuals' vocabulary and by empowering everyone to organize a

collection their own way. In tagging, keywords can be chosen freely and are treated identically with no hierarchical background.

These user-added keywords are the basis organizational objects for the emergent folksonomies. While by tradition, metadata was created primarily by experts following stringent taxonomies and pre-specified controlled vocabulary, the categories based on Web user-created metadata are more flexible [Orio, 2010]. To return to the library analogy (see sect. 2.1), experts agree on the usage of specific metadata (i.e. taxonomies) to annotate content. Web users, however, are not bound by such agreements. Hence, in contrast to taxonomies, which are hierarchical and exclusive, tags are neither exclusive nor hierarchical. The three entities tags, users, and resources constitute what is called a folksonomy [Smith, 2008].

UGC and sincerity are coevally the folksonomies' advantage and disadvantage. Its simplicity and low entry barriers comfort people to actively participate in tagging and thus inflicting metadata to the Web [Hächler, 2010]. It is a very facile process that takes no ancillary capabilities because each user can use his own vocabulary. While the traditional and professional creation of metadata is time and effort consuming, folksonomies can keep up with the immense amounts of new content timelessly being created on the Web. They allow quick adaption of new terms when traditional vocabulary is missing. However, there are limitations resulting from the democratic way of labeling Web content. Tagging's nature, and therefore that of folksonomies, is fundamentally chaotic, prevalently giving rise to problems of imprecision and ambiguity because there is just no predefined vocabulary to be used.

All in all, folksonomies are transforming the creation of metadata for resources from an isolated professional activity into a shared, communicative activity by the users. This shift is of dual nature and also causes some difficulty: On one hand, tagging is a great system for individual organizations, at the same time there is an inherent compulsion to share in order to generate folksonomies and to reveal the full and useful power of the system for the user. However the dualisms should be carefully considered. Folksonomies are aggregated through vast amounts of metadata created by the users. The fundamental difference to traditional classification schemes lies in the reduced complexity. Some organizations use folksonomies to let their employees easier manage their own Intranet hyperlinks.

A further possibility for organizing information and knowledge instead of metadata, are wikis. A wiki is a website that allows the creation and editing of interlinked websites via a Web browser using a simplified markup-language.

2.2.4 WIKIS

A wiki (e.g. Wikipedia or Wikitravel) is a system that allows one or more people to build up a knowledge body of interlinked webpages, using a pro-

cess of creating and editing pages. With wikis, anybody can contribute equally to a joint online publication. Wikis are rooted on the convention that contributors can straightly post whatever they know about a topic for others to approve, clarify, add to, or revise rather than all content is being accepted by a Web administrator as occurs with conventional websites. Centralized production and top-down techniques of knowledge sharing are being pushed aside by the belief and the new concept that everyone together is smarter than one alone. This concept is known as collective intelligence, a shared group intelligence that emerges from the collaboration and competition of many individuals [Malone, 2006]. This concept can also be found more or less in every social media element. For example, the previously introduced social bookmarking platforms, where users are collaboratively organizing Web content in a more meaningful way as only experts can do. The times when knowledge needed to be vouched for, authorized and approved by experts before it could hit the broad audience seem in some areas to fade out evermore. Wikis are characterized by arising properties in the media.

The aim of the wiki is to establish collective knowledge. Therefore, they highlight the participation, the contribution and collaboration of the users. The keyword for wikis is easy. Anyone can edit and contribute to a wiki, demanding it to be easy to handle. Wikis usually have an editing link, through which anyone can start writing and adding to the content. They tightly pursue the open-source software ideal, which implies that the quality of the collectively produced product is more crucial than owning the idea or the code. Wikis can play havoc with the conventional ideas of copyright and intellectual property [Richardson, 2010].

Whereas blogs and microblogs are good for discussions, wikis are not as optimal for carrying out discussions about conceptions. Most conversations, concerning which article and what posts need to be altered in what way, may happen parallel to it. However, this process demonstrates how the collaboration is brought to light and correspond with the ideal of open-source knowledge gathering [Hächler, 2010].

In some organizations wikis are tools successfully used to manage the organization's internal information and knowledge. Moreover, they can be used in organizational context also for project management. Through wikis newcomers to the organization or a project can gain fast access to (relevant) information [Fuchs, 2010].

The last crucial elements of the Social Web are social networks, presented in the next section. These networks are social structures made up of individuals (or organizations) called nodes, which are connected by one or more specific types of interdependency.

2.2.5 SOCIAL NETWORKS

Social networks (e.g. Facebook or Google+) have become a significant medium for self-expression and identity generation. A vital characteristic of the present culture is the raising of individualism and the elevation of the individual experience as a guarantor of truth. The individual has increasingly evolved to the heart of social, economic and technological order. Besides, society seems to become evermore formed by an expressive culture, blurring the differentiation between private and public. Stepwise sociality is taking place in the virtual world and likewise intimacies are being carried into the virtual world. In such an environment it then becomes a critical factor as to how individuals cultivate and negotiate their own identity. Each person is responsible for drawing a suitable persona. Therefore, the resulting persona is as intimate as its public.

A definition of social network sites stems from [Humphreys, 2009], who calls them "*Web-based services that allow individuals to* […] *construct a public or semi-public profile within a bounded system,* […] *articulate a list of other users with whom they share a connection, and* […] *view and traverse their list of connections and those made by others within the system.*" For [Levinson, 2009], this kind of social media has the very purpose of building and developing social networks, and thus enabling people to connect for whatever proposition. Social networks provide their members a platform from which they can engage in a wide-ranging variety of activities such as private messaging, bulletins or group messaging, blogging, posting of photos, videos or music, as well as instant messaging and groups devoted to common interests. Additionally the platform also allows tagging content (e.g. photos, links, etc.). The topic of this social medium is to bring people together. Groups and similar online activities such as forums or message boards are then fundamental elements of online life. Groups share and discuss links, texts, photos and videos. This interconnectedness connote that the most relevant component of any social network is the friend, follower or contact. They can be known solely virtually or be known in real life also.

There is one clandestine dimension to all the self-productions through social media and that is the persistence of everything once it has been published [Hächler, 2010]. Publications can have an effect which might have been unintended or unforeseen at the time of publication. Users leave data trails which are being collected incessantly. This collection happens automatically, invisibly and mostly involuntarily. Users might feel that they own the social media because of the extraordinary power of production and self-projection it furnishes.

In business environments social networks can be used to promote products or brands. Yet, social networks can extend the outreach of an organization not only to promote, but also to gather information and knowledge—also through personal employees' connections.

2.3 THE VISION OF THE SEMANTIC WEB

Although the Social Web offers an easy mode to share information, work collaboratively and maintain relationships, the capability to read, understand, and process content is limited only to humans. Computers have difficulties handling documents with natural language content, not to mention handling them automatically. Information as preferred by humans is hard to find too, as the precision of search results is low. Searching for information rests upon identifying words within websites and matching them. For example, if someone is searching for Apple, normally a Web search engine as Google[17] is consulted. Yet, based on the search term, the Web search engine will unfortunately refer to the multinational organization Apple Inc. and the fruit (and others) with no opportunity to discriminate, if there is no additional contextual information available. Acronyms can similarly be a problem, such as ANT being the insect, Apaches Another Neat Tool for automating software build processes or the acronym for Actor Network Theory. Using more than one search term can help in finding websites more precisely, but there is no control over synonyms. The same resources can be described in different ways. Distorted search results may in turn prompt frustrated Web surfers to crowdsource their questions to trusted branches of their social networks, for example. The idea is that friends and relations can steer someone toward a good answer much more reliable than today's Web search engines. Nevertheless, the Semantic Web promises a remedy.

Up until now Web search engines had difficulties to put a search query into context with the implicit user's need. It would be of great help if computers could assist users and ease the load by having computer-understandable semantics at their disposal in order to understand findings like humans can and thereby avail oneself of natural language Web content. This is exactly what the Semantic Web is conceived to establish. [Berners-Lee et al., 2001] sketched the idea as follows: *"The Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computer and people to work in cooperation."* This section is now concerned with the technological aspects of the Semantic Web. The extension of the present Web is made up of metadata that delineate the semantics of the content of websites. The Greek word semantic stands for the meaning of, and consequently the Semantic Web represents a Web that is able to understand its content; this is accomplished mainly by embedding further meaning to the Web. The idea of a Semantic Web implicates a move from unstructured websites (e.g. without or only with sparse computer-understandable metadata) to structured ones that cannot only be understood by humans. The semantic vocabulary is based on concepts that are defined in ontologies.

The term ontology (from the Greek words on meaning being and logos meaning to reason) was originally coined in philosophy to denote the theory or study of being as such. The use of the term ontology in computer science has a more practical meaning than its use in philosophy. The study of meta-

physics is not in the foreground in computer science, but rather what properties a computer must have to enable it to process data that is being questioned within a certain domain of discourse. Ontologies are artifacts of objects and their ties. Hence ontologies provide criteria for distinguishing various types of objects (e.g. concrete and abstract, existent and non-existent, real and ideal, independent and dependent, etc.) and their ties (relations, dependences and predication). Within computer science, the term stands for a design model for specifying the world that consists of a set of types, relationships and properties. What is provided precisely can deviate, but these properties are fundamentals of every ontology.

According to [Gruber, 1993], an ontology is a *"formal, explicit specification of a shared conceptualization"*. There is an expectation that the model bear analogy to the real world as well; however, it definitely offers a common terminology that can be used to model a domain. A domain is the type of objects and concepts that exist, and their properties and relations. In literature there is furthermore a distinction between strong and weak ontologies, whereas in this thesis only weak ontologies are used. A weak ontology is one that is not sufficiently as rigorous as a strong one and therefore allows computers to insert new details without an intervention by humans. In addition, a weak ontology converges with Description Logic (DL) and other subfields in which automatic reasoning is known to be possible.

Ontologies provide a vocabulary of terms in a given field that are needed to itemize the meaning of the annotations added to websites. Consider, for example, the domain of `organization`. This domain contains concepts such as `name`, `project`, `employee`, `store`, etc. Each `store` has for example a property `isInRegion` and a relationship `belongsToOrganization`; the latter links a `store` to the concept `organization`. Likewise an `employee` is a `person` such as Tim Cook, the Chief Executive Officer (`CEO`) of Apple Inc. It is possible to say that the ontology provides concepts, properties, and relationships with well-defined meanings such that the business information of websites can be described and annotated by relying on the elements of the ontology. So ontologies are designed to be understandable by computers as part of the Semantic Web.

In figure 2.2 an abstract of the Semantic Web Stack (aka Semantic Web Cake or Semantic Web Layer Cake) is presented; it is composed as a hierarchy of languages, where each layer engages capabilities of the underlying layer. The Stack is still evolving as the layers are concretized. Accordingly in this figure only the layers used within this thesis are highlighted.

The bottom layers of the Stack contain technologies such as Unicode, Uniform Resource Identifier (URI), and eXtensible Markup Language (XML). These technologies are well known from the earlier Web and, without modification, provide a root for the Semantic Web. The presented middle layers

enclose technologies standardized by W3C to enable building Semantic Web applications.

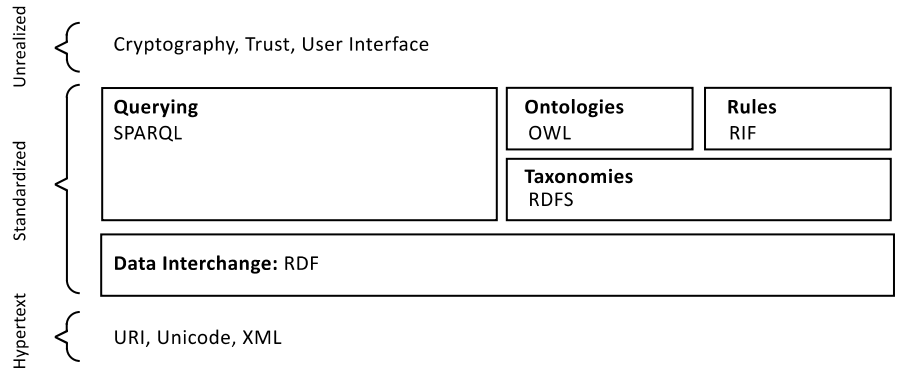

Figure 2.2: Semantic Web Stack.

In the following sections, the individual layers of the Stack will be introduced. Section 2.3.1 begins with the Resource Description Framework (RDF); a framework for creating statements in the form of triples (i.e. subject, predicate, and object). RDF presents information about resources in the form of a graph. Section 2.3.2 introduces RDF Schema (RDFS). RDFS provides basic vocabulary for RDF. Using RDFS it is possible to create hierarchies of classes and properties. The Web Ontology Language (OWL) extends RDFS by adding more advanced constructs to describe semantics of RDF statements. It allows stating additional constraints, such as cardinality, restrictions of values, or characteristics of properties such as transitivity. It adds expressiveness to the Semantic Web, described in section 2.3.3. The Rule Interchange Format (RIF) brings support for rules and is presented in section 2.3.4. This is important to allow describing relations that cannot be directly described using DL as used in OWL, for example. Lastly, the SPARQL Protocol and RDF Query Language (SPARQL) is envisaged in section 2.3.5. This is a RDF protocol and query language—it can be used to query any RDF-based data (i.e. including statements involving RDFS and OWL). Hence, SPARQL is necessary to retrieve information from the Semantic Web.

At last, the top layers contain technologies that are not yet standardized or hold just concepts, such as cryptography or a trust layer, which should be implemented in order to realize the Semantic Web. Within this PhD thesis, these layers are not further explained.

2.3.1 Resource Description Framework

RDF is used to represent entities, referred to by their unique identifiers or URIs, and a binary relationship among those entities. RDF is made of two parts: The data model specification and serialization syntax. The data model

definition is the core of the specification, and the syntax is essential to convey RDF data in the Web.

Two entities and their binary relationship are termed a statement or a triple. Shown graphically, the source of the relationship is termed the subject of the statement, the labeled arc itself is the predicate (or property) of that statement, and the destination of the relationship is called the object of that statement. The data model of RDF distinguishes among entities (or resources), which have a unique URI identifier, and literals, which are solely strings. The subject and the predicate of a statement are always resources, while the object can be a resource or a literal. In RDF diagrams, resources are drawn as ovals, and literals as boxes. An illustration of a statement is given in figure 2.3. This statement is adapted from [Breslin et al., 2009].

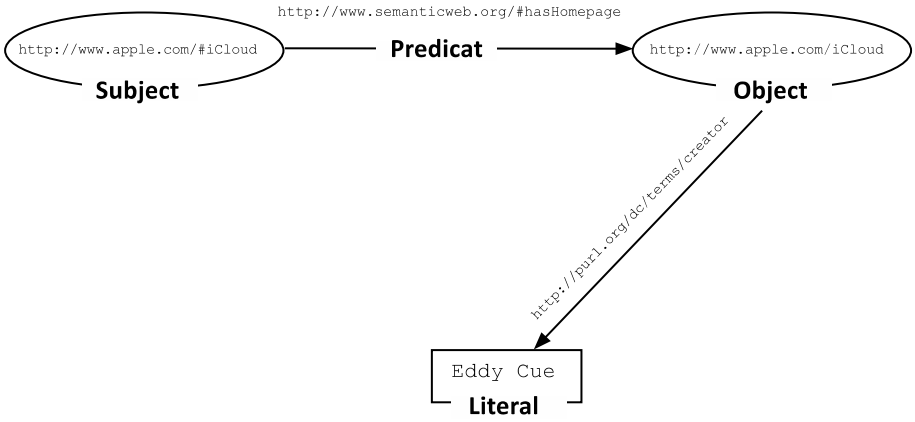

Figure 2.3: Simple RDF Graph.

This can be read as that the resource `http://www.apple.com/#iCloud` has a homepage, which is `http://www.apple.com/iCloud`. At first glance it may look odd that predicates are resources too, and thus have a URI as a label. However, to prevent confusion it is essential to give the predicate a unique identifier. Simply `hasHomepage` would not meet the requirements, as different vocabularies might state different descriptions of the predicate with different meanings. The property `http://purl.org/dc/terms/creator` with value "`Eddy Cue`" (a literal) has been added to the graph to indicate that Eddy Cue (i.e. Apple's responsible for Web software and services who oversees iCloud[18] service) has created the homepage.

For a more functional data representation, further vocabularies and conventions should be established. For instance, predicate URI's are usual shortened by employing the XML-namespace syntax. Instead to write the full URI `http://www.SemanticWeb.org/#hasHomepage`, the namespace form `sw:hasHomepage` is employed with the hypothesis that the substitution of the namespace prefix `sw` with `http://www.SemanticWeb.org/#` is

defined. In addition, the namespace prefix `rdf` is often used to refer to the specification declaring how metadata should be created according to the RDF model and syntax [Lassila & Swick, 1999]. In this case, the `rdf` prefix would be extended to the Uniform Resource Locator (URL) of the RDF-specific vocabulary `http://www.w3.org/1999/02/22-rdf-syntax-ns#`. The same goes for RDFS model and syntax as well (see sect. 2.3.2).

The RDF specification suggests two standard ways to serialize RDF data in XML: A shortened and a standard syntax. Both serialization possibilities use the XML namespace mechanisms to reduce URIs as already presented [Bray et al., 1999]. Another option for serializing RDF is the annotation of HTML5 documents with RDF in Attributes (abbreviated as RDFa). RDFa allows including semantics in HyperText Markup Language (HTML), so that the data can be mapped to RDF and objects can be identified by URIs [Breslin et al., 2009; Pilgrim, 2010]. This approach virtually bridges the gap between the Semantic Web for humans and for computers since a single document with RDFa can cover information for both. This also circumvents the repetition of information between a HTML and an RDF/XML document.

As presented in this section, RDF is used as a general method for conceptual description or modeling of information that is implemented in Web resources, using a variety of syntax formats. These syntax formats can be found in [Hitzler et al., 2010].

2.3.2 RDF SCHEMA

The objective of the RDFS specification is to determine the primitives needed to describe classes, instances and relationships [Brickeley & Guha, 2004]. RDFS is an RDF application, defined in RDF itself. The defined vocabulary is similar to the usual modeling primitives available in frame-based languages (where domain entities are modeled as frames that have a set of appropriate slots or properties). In this section, the vocabulary used in the stated examples is defined by RDFS. The prefix `rdfs` is thus an acronym for `http://www.w3.org/2000/01/rdf-schema#`, the RDFS namespace identifier.

Figure 2.4 depicts an RDFS-based ontology, defining the class `sw:Project` and two properties `sw:hasHomepage` and `sw:hasMember`. The class node is defined by typing the node with the resource `rdfs:Class` that represents a metaclass in RDFS. `sw:Project` is also defined as a subclass of `rdfs:Resource`, which is in the class hierarchy the most general class defined by RDFS. The `rdfs:subClassOf` property is defined as transitive and `rdfs:Literal` represents the class of XML literal values (e.g. strings and integers). The presented RDFS-based ontology is adapted from [Breslin et al., 2009].

Properties are defined with the resource `rdf:Property`, which is the class of all properties. `rdf:type` is an instance of `rdf:Property` used to state that a resource is an instance of a class. Furthermore, the domain and range of a property can be constrained by using the properties `rdfs:range` and `rdfs:domain` to define value restrictions on properties. For example, the property `sw:hasHomepage` has the domain `sw:Project` and a range `rdfs:Resource` (which is compliant with the use of `sw:hasHomepage` in figure 2.3). Using these definitions, RDF data can be tested with conformance in relation to a particular RDFS specification. RDFS defines even more modeling primitives, which could be found by [Hitzler et al., 2010].

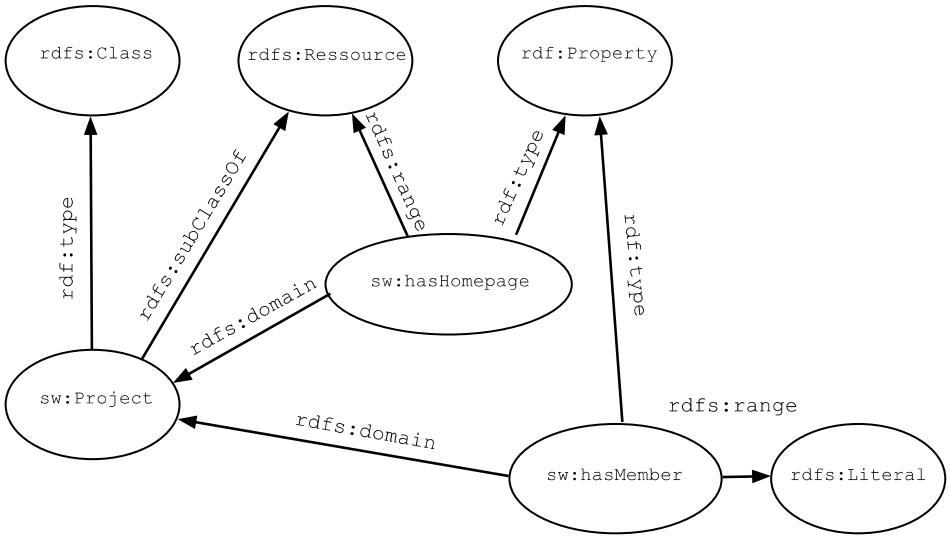

Figure 2.4: RDFS-Based Ontology.

It is usual for ontologies to refer only to the ontology schema (aka the ontology model or ontology meta-model). As introduced, an ontology is simply a specification of conceptualization without naming instances. If instances are annotated by ontology tags and modeled as ontology, then the talk is of a knowledge base (or an ontology-based knowledge base). Thus, a knowledge base is a collection of instances of the concepts defined in the ontology, and the ontology specifies the structure of the knowledge stored in the knowledge base [Kaufmann, 2007]. Further terminology involves the terms ABox and TBox, which are employed to define two statement types in ontologies and knowledge bases: TBox statements describe the controlled vocabulary or the set of classes and properties of an ontology. ABox statements are TBox-compliant statements regarding the vocabulary that describe instances. For instance, a specific apple tree is an individual for the concept of tree, while it can be stated that trees as a concept are material beings that have to be positioned on some location it is possible to state the specific location that an apple tree takes at some specific time. Together, all ABox and TBox statements make up the ontology-based knowledge base.

2.3.3 WEB ONTOLOGY LANGUAGE

The expressiveness of RDFS is rather restricted. For instance, it cannot be used to define that a property is symmetric (e.g. `:isNeighbourOf`) or transitive (e.g. `:locatedIn`). This restricted expressiveness resulted in the necessity for a more powerful ontology modeling language—in particular, one that permitted widened computer interpretability of Web content. This led to OWL, which allows modelers to use an expressive formalism to define diverse logical concepts, and relations in ontologies to annotate Web content [McGuinness & van Harmelen, 2004]. Computers can then use the strengthened content in order to assist humans in various tasks. As such, OWL performs the requirement for an ontology language that can formally describe the meaning of terminology on websites. If computers are anticipated to perform useful reasoning tasks on these Web documents, the language must transcend the semantics of RDFS. OWL has been designed to meet this necessity.

In the same fashion of RDFS, OWL can be utilized to explicitly represent the meaning of terms in vocabularies and the relationships between those terms. The ontologies are used by applications that need to process the content of information instead of just presenting it. OWL provides a more substantial vocabulary in the context of formal semantics than RDFS by permitting additional modeling primitives that result in an enhanced expressiveness for characterizing properties and classes. A complete list of OWL syntax can be found in [Hitzler et al., 2010]. OWL offers three sublanguages with varying degrees of expressiveness. These are OWL Full, OWL DL, and OWL Lite (ordered by descending expressiveness). Each of these sublanguages is a syntactic extension of its simpler predecessor.

- *OWL Full*: This is the complete OWL language without any limitations and complete with maximum expressiveness, but lacking any computational guarantee. All language constructs can be used in any combination as long as the result is valid RDF.

- *OWL DL*: This limits the expressive power of OWL Full (and increases the expressive power of OWL Light). It offers all OWL constructs with certain limitations such as type separation. For example, every resource can only be a class, a property, or an individual. This means that a class cannot simultaneously be an individual. OWL DL is intended for people who want maximum expressiveness, but retain computational completeness (all conclusions can be computed) and decidability (all computations will finish in finite time).

- *OWL Light*: This further restricts the expressive power of OWL DL. It also offers hierarchies of classes and properties, and simple constraints enable the modeling of thesauri and simple ontologies. Limitations are imposed on how classes are related to each other.

The OWL family contains several species, serializations, syntaxes and specifications with similar names. This might be unclear unless a consistent approach is implemented. OWL and OWL2 are used to refer to the 2004 and 2009 specifications, respectively. Full species names will be used, including specification version. When referring more generally, the OWL family is used.

Specification in this context means an explicit representation by some syntactic means. Most approaches to ontology modeling agree on the following primitives for representation purposes: Firstly, there must be a distinction between classes and instances, where classes are interpreted as a set of instances. Classes may be partially ordered using the binary relationship `:subClassOf`, which can be interpreted as a subset relationship between two classes. The fact that an object is an element of a certain class is usually denoted with a binary relationship such as `:type`. Consider, a combination of an ontology together with a set of instances of classes constitutes a knowledge base (see sect. 2.3.2). Secondly, a set of properties (also called attributes or slots) is required. Slots are binary relationships defined by classes, which usually have a certain domain and a range. Slots might be used to check if a certain set of instances with slots is valid with respect to a certain ontology.

Another important fact to keep in mind regarding these languages and the Semantic Web in general is that they refer to what is termed an Open-World Assumption (OWA). Consequently, if something is not defined, nothing can be anticipated about it. By way of example, if no triples bring up that `:Heather :isMarriedWith :Jony`, and if someone asks if Heather and Jony are married, the answer will not be no but rather unknown, as there are not enough facts to answer that query.

2.3.4 Rule Interchange Format

RIF is designed as a general framework for interchanging rules of various types. A possibility to converge this ambitious target is to start with a least common denominator of a set of rule languages. Such a shared core language helps to highlight the commonalities of various formalisms and can be a foundation for leading the expansion toward more meaningful languages. Although initially meant by many as a rules layer for the Semantic Web, indeed the design of RIF is based on observations that there are many rules languages in existence, and what is really needed is to exchange rules between them [Hitzler et al., 2010].

A rule is arguably one of the elementary concepts in computer science: It is collected from an `IF-THEN` construct. `IF` some condition that is testable in a dataset holds, `THEN` the conclusion is processed. Deriving anything from its DL roots, rule systems employ a notion of predicates that hold or not of some data objects. For instance, the above mentioned fact that Heather Pegg

is married with Jony Ive (i.e. Apple's designer behind iPad, iPhone, and iPod) might be expressed with predicates as `:married(:Heather,:Jony)`. `:married` is a predicate that can be claimed to hold between `:Heather` and `:Jony`. Adding the notion of variables, a rule could be:

```
IF :married(?x,?y) THEN :loves(?x,?y)
```

Thereby it is expected that for every pair of `?x` and `?y` (e.g. `:Heather` and `:Jony`) for which the `:married` predicate holds, a computer that could understand this rule would deduce that the `:loves` predicate holds for that pair too.

Rules are an elementary type of encoding knowledge, and are a rigorous simplification of DL for which it is comparatively straightforward to implement reasoning engines (e.g. Bossam, FaCT++, HermiT, Pellet, RacerPro, etc.) that can handle the conditions and draw correct conclusions. A rule system is an implementation of a certain syntax and semantics of rules, which may expand the elementary notion from above to enclose existential quantification, disjunction, logical conjunction, negation, functions, non-monotonicity, and many other features. The standard RIF dialects are RIF-Core, Basic Logic Dialect (BLD), and Production Rule Dialect (PRD) [Hitzler et al., 2010]:

- *RIF Core*: This dialect is supported by a large class of rule-based systems and is defined as a restriction of the more expressive BLD, but it can similarly be considered as a sublanguage of PRD. In this meaning, RIF-Core is effectively the fundamental core of the rule languages envisioned by RIF. Semantically, RIF-Core is closely allied to DL programming without function symbols or any pattern of negation.

- *RIF BLD*: This dialect adds features to the Core dialect that are not directly available such as logic functions, equality in the `THEN`-part and named arguments. It corresponds to positive logic programs without functions or negations and has a model-theoretic semantics. From a semantic point of view, the main variation between RIF-Core and RIF-BLD is that the latter also facilitates the employ of function symbols.

- *RIF PRD*: This dialect can be used to model production rules. Features that are predominantly in PRD but not in BLD include negation and retraction of facts. These rules are order dependent, hence conflict resolution strategies are required when multiple rules fire. The PRD specification defines such a resolution structure based on forward-chaining reasoning. RIF-PRD has an operational semantics, whereas the condition formulas also have a model-theoretic semantics.

2.3.5 SPARQL PROTOCOL AND RDF QUERY LANGUAGE

RDF(S) and OWL are effective languages for representing ontologies and metadata on the Semantic Web. However, as soon as this metadata has been published, query languages are needed to draw full benefit of it. SPARQL satisfies this aim and allocates a query language and a protocol for RDF data on the Semantic Web. By providing HyperText Transfer Protocol (HTTP) bindings for it, as well as normalized serialization of the results—in XML or JavaScript Object Notation (JSON) —it can be efficiently used to provide open access to RDF databases.

SPARQL can be thought of as the Structured Query Language (SQL) of the Semantic Web, and offers a powerful method to query RDF triples and graphs. As RDF data is elucidated as a graph, SPARQL is therefore a graph-querying language, which means that the approach is distinct from SQL where one is concerned with tables and rows. Moreover, SPARQL provides extensibility within the query patterns (based on the RDF graph model itself) and therefore advanced querying capabilities on the basis of this graph representation.

SPARQL can be used to query standalone RDF files as well as sets of RDF files, either loaded in memory by the SPARQL query engine or through the utilization of a SPARQL-conformal triplestore (a storage system for RDF data). Therefore, there is presently a need to know which files must be queried before running a query, which is a hitch in some cases and can be regarded as a hurdle to be overcome. However, in addition to distributed SPARQL query engines in order to dynamically identify which RDF sources should be considered when querying information, four different approaches are normally used: `ASK`, to postulate a simple true-or-false result for a query on a SPARQL endpoint; `CONSTRUCT`, to abstract information from the endpoint and transform the results into valid RDF; `DESCRIBE`, to extract an RDF graph from the endpoint; and `SELECT`, to extract raw values from an endpoint. Note that each of these query forms takes a `WHERE` block to restrict the query although in the case of the `DESCRIBE` query the `WHERE` is optional [Breslin et al., 2009; Hitzler et al., 2010; Seaborne et al., 2008].

Query patterns generate an unordered collection of solutions, with each solution being a partial function from variables to RDF terms. These solutions are treated as a sequence with no specific order. Sequence modifiers can then be applied to create an order: `DISTINCT` ensures that the solutions in the sequence are unique; `LIMIT` restricts the number of solutions; `OFFSET` controls where the solutions start from in the overall sequence; and `ORDER` puts the solution in a specific order. `FILTER`s are other conditions in a query that restrict a set of matching results. Contrasting graph patterns, `FILTER`s are not only based on RDF, but may cover further requirements.

While SPARQL is obviously a key component of the Semantic Web, it has some limits. At the time of writing, SPARQL does not provide any aggregate function, hence implying a need to use external languages (e.g. PHP Hypertext Preprocessor (PHP) or JavaScript) to run aggregations, which can make the adoption of RDF technologies complicated in some cases. Furthermore, SPARQL is a read-only language, in that it does not allow one to add or modify RDF statements. Likewise also vagueness is not adequate supported by SPARQL so far [Stoilos et al., 2005b]. Thus, long ways round have to be taken to overcome a vague query [Cheng et al., 2010]. This is occasionally quite an obstacle, which is why within this PhD project `FILTER`s are used for fuzzy querying (see chap. 7).

2.4 Towards a Social Semantic Web

The last years have shown immense undertakings for the definition of the foundational standards supporting data interchange and interoperation. A number of Semantic Web technologies have attained broad deployment. Often these technologies are composed of ontologies, which share a property: They are small and vertical; in other words, they are member of numerous domains. Each horizontal domain (e.g. `:organization`) would typically reuse a wide range of these vertical ontologies, and when deployed the ontologies allow interacting with each other.

In a most helpful starting point, the Semantic Web attempts to make social websites interoperable by providing standards to support data interchange and interoperation between applications, empowering individuals and communities to partake in the construction of shared interoperable information. This adaption of the Semantic Web to the Social Web gives rise to either a social Semantic Web (i.e. more top-down driven) or a semantic Social Web (i.e. more bottom-up driven), summarized as Social Semantic Web that is an innovative Web of interlinked and semantically-rich knowledge. This vision of the Web is made of interlinked documents, data, and even applications created by the users themselves as a consequence of all kinds of social interactions, and it is depending on computer-readable formats so that it can be used for purposes that the actual state of the Social Web cannot accomplish without difficulty [Breslin et al., 2009].

Adapted from [Blumauer & Pellegrini, 2009], figure 2.5 pictures prototypically the different paradigms to evolve a Social Semantic Web. Thereby the two different kinds of indexing (manual vs. automatic) are represented by the horizontal axis and the two different kind of knowledge organization (expert vs. community) by the vertical axis.

As previously discussed, the libraries were the first to use expert-based manual knowledge organization. The ideas described were either motivated by a community-based knowledge organization or by an automatic indexing of the data by computers. Nowadays the Social Semantic Web can connect

these ideas and generate a symbiosis of collective intelligence between humans and computers. Through adding social features to the Semantic Web the social Semantic Web emerges, while the injection of computer-understandable semantics to the Social Web yields the semantic Social Web; within this PhD thesis the focus is rather on the latter. However, yet, in literature no distinction can be found and so, with a few exceptions, this thesis adheres to the common used term Social Semantic Web.

With the Social Semantic Web it is now feasible to harness the intelligence of vast numbers of people, connected in very different ways and on a considerably larger scale than has ever been imaginable before. As forms of collective intelligence grow in importance, as seen with crowdsourcing projects, the value of socially aware individuals is going to arise as well [Saenz, 2011].

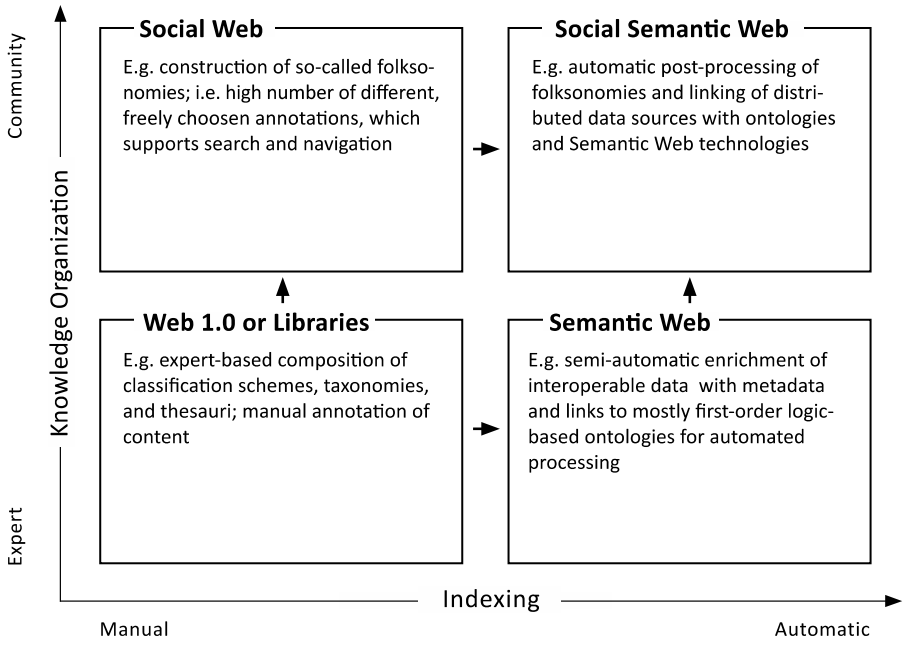

Figure 2.5: Development of the Social Semantic Web.

Subsequently, by a clever linking of human and computers strengths, this can lead to an enhanced collective intelligence. The combination of the Social Web and Semantic Web can lead to something greater than the sum of its parts, where the Social Web elements can be interconnected with Semantic Web technologies, and Semantic Web elements are enhanced with the wealth of knowledge inherent in UGC. According to [Breslin et al., 2009], this goes hand-in-hand with solving the chicken-and-egg problem of the Semantic Web (i.e. it is difficult to create useful Semantic Web applications without the data to power them, and it is difficult too, to produce semantically-rich data without the interesting applications themselves). Since the Social Web entails such semantically-rich content, interesting applications powered by

Semantic Web technologies can be created straightway [Blumauer & Pellegrini, 2009]. In terms of the increasing integration of mobile devices and everyday objects into the Web, this is also highly favored.

For example, Social Web users are already bringing semantically-rich annotations through folksonomies into being. This PhD thesis' intention for the Semantic Web is to amend the bottom-up attempt of the Social Web in a top-down manner as [Cardoso, 2007] suggests. The fundamental aim is a stronger knowledge representation, as can be achieved with folksonomies solely, for example. Fuzziness can overcome the gap between folksonomies and ontologies because fuzziness corresponds to the way in which humans think and it is, thus, suitable for characterizing vague information and helps to more efficiently handle real-world complexities [Meier et al., 2008]. One possible way to use these advantages is through fuzzy clustering algorithms, which allow modeling of the uncertainty associated with vagueness and imprecision through mathematical models [de Oliveira & Pedrycz, 2007; Bezdek et al., 2008; Miyamoto et al., 2008]. As a crucial part of this PhD thesis, folksonomies will be converted to computer-understandable ontologies adapted from fuzzy clustering algorithms. With this in mind, the next chapter introduces into the fundamentals of fuzzy clustering methods. In the end, the completely computer-produced ontology will be used to enhance online reputation analysis in the Social Semantic Web.

2.5 Further Readings

A book worth reading about libraries in ancient world is [Too, 2010]. The connection to the Web and its history is given by [Banks, 2008], [Hafner & Lyon, 2008], and by [Portmann, 2008]. The development of the Web to a Social Web can be extracted from [Alby, 2008] and [Ebersbach et al., 2010]. In these books also descriptions of the various Social Web applications such as blogs, folksonomies, microblogs, social networks and wikis can be obtained. [Portmann & Hutter, 2011] explain the interaction of these Social Web applications, and [Jenkins, 2008; Manovich, 2001; McLuhan & Nevitt, 1972; Toffler, 1980] broadly explain the shift from traditional to interactive media; thereby their focus is on media sciences.

Information about the Semantic Web can by now be found in numerous books. Thereby [Antoniou & van Harmelen, 2008] and [Allemang & Hendler, 2008]'s focus is on practical usability of the Semantic Web, whereby [Hitzler et al., 2008] and [Hitzler et al., 2010] condense more the theoretical potential of the Semantic Web, but likewise include various examples of the power of its technologies.

The Social Semantic Web and its possibilities are explained in [Breslin et al., 2009] or by [Blumauer & Pellegrini, 2009]. Thereby also other practical applications bridging the gap between Social and Semantic Web are presented. Such applications are, for example, Dublin Core[19], DBpedia[20], Semantic Media Wiki[21], Friend-Of-A-Friend (FOAF) protocol[22], Description Of A Project

(DOAP)[23] or the Semantically-Interlinked Online Communities (SIOC)[24]. Explanatory notes with concerning illustrations on these tools can be found in [Breslin et al., 2009; Blumauer & Pellegrini, 2009; Hitzler et al., 2010].

3 FUNDAMENTALS OF FUZZY CLUSTERING METHODS

"As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality."

—Albert Einstein

Most of the conventional tools for formal modeling, reasoning, and computing are hard, deterministic, and precise. Thereby hard implies unambiguity that is, yes-or-no rather than more-or-less. In traditional bivalent logic, for example, a statement can be true or false—and nothing in-between. Precision assumes that parameter of a model typifies precisely the features of a real system that has been modeled. Usually, precision also implies that a model is doubtless, that is, that it covers no ambiguities [Zimmermann, 2001].

In the end, certainty signifies that the characteristics and patterns of a model are ultimately known, and that there are no disbeliefs about their merits or their occurrence. Taking into account that a model is formal, that is, if it does not pretend to model reality, then the modeler can freely pick its characteristics. If, however, the model argues for factuality, that is, if conclusions drawn from these models have a bearing on reality and are supposed to model reality, then the modeling language has to be suited to model the characteristics of the situation reasonable [Zimmermann, 2001]. The usefulness of mathematics as modeling language is undisputed. However, there are limits to the usefulness and the possibility of using classical mathematical language to model particular systems and phenomena in information systems, computer and social science: Real situations are frequently ambiguous, and cannot be described precisely. Furthermore a complete description of a real system often requires far more detailed data than human modelers could ever realize, process, and understand simultaneously.

Fuzziness, in turn, can be found in many areas of daily life, such as engineering [Pedrycz & Gomide, 2007], manufacturing [Venkata Rao, 2010] or medicine [Barro & Marin, 2010]. It occurs most in areas in which human judgment, analysis, and decision are important. Reasons for fuzziness are that most of humans' everyday communication use natural languages and a good part of thinking are done in it. In these natural languages, the meaning of words is very often vague. The meaning of a word might even be well defined, but when using the word as the label for a set, the boundaries within which objects do or do not belong to a set become vague or fuzzy. Examples are words as birds (how about penguins?) or red apples (how red is red?), but also terms as beautiful design, lightweight computer, and large organization. Even for a single person it may not be possible to give a precise and clear answer as belonging to a set (e.g. red apples) is often not sharp but fuzzy [Zimmermann, 2001]. Fuzziness incorporates a partial matching expressed in natural language by the verbalisms very, slightly, and more or less, etc.

This chapter also contains to a great deal of theory. In the course of this, it should come across as additional to the previous one. By cluster analysis (or clustering) Social Web content can be extended to Social Semantic Web content. Section 3.1 of this chapter gives a brief introduction into clustering that is exactly the mentioned assignment of an object to a set. To this end, section 3.2 presents clustering for object data and showcases its general principles. Subsequently, in section 3.3, the general (hard) approach of clustering is enhanced by a more flexible (fuzzy) approach; thereby the main concepts and mathematical notions of fuzzy logic and fuzzy set theory are presented and as a result fuzzy clustering is explained in more detail. Section 3.4 applies fuzzy clustering to the Social Semantic Web. Last but not least, section 3.5 concludes this chapter with further readings.

3.1 Introduction to Cluster Analysis

Cluster analysis (or clustering) has become a widely accepted synonym for a wide range of activities of exploratory data analysis and model development. Areas such as data mining, Information Retrieval (IR), image analysis, pattern recognition, modeling, and bioinformatics are concrete examples of many attempts that utilize the concepts and algorithms of clustering treated as essential tools for problem formulation, solutions development, and interpretation mechanism. The aim of clustering is the organization of the dataset into homogenous or natural classes, in a way, which ensures that objects within a class are similar to one another. In this way clustering weeds through a dataset with the aims at decomposing this into subclasses based on similarity. Thereby the dataset is divided in such a way that objects (e.g. example cases, elements, individuals or observations) belonging to the same class (or cluster) are as similar as possible, whereas objects belonging to different classes (or clusters) are as dissimilar as possible [Miyamoto et al., 2008].

There is some confusion about the use of the terms classification and clustering. Classification, whose task is to assign objects to clusters or groups on the basis of measurements of the objects, is more general and can be divided into supervised and unsupervised classification. Supervised classification (or discrimination) seeks to create a classifier for the classification of future observations, starting from a dataset of labeled objects (a training or learning set), whereas unsupervised classification (or clustering) seeks to discover groups from the dataset [Govaert, 2009; Bishop, 2007].

This PhD thesis is restricted to unsupervised classification methods. However, on one or another part of this PhD project an enhancement by a supervised classification is feasible. In each case this is mentioned causally as an additional option. The particular terminology depends on the underlying field: Taxonomy, for example, is the branch of biology concerned with the classification of organisms into groups. In machine learning, clustering is known as unsupervised learning; in marketing the talk is about segmentation; and in linguistics, the most frequently used term for clustering is typology [Govaert, 2009].

Hence clustering is primarily a tool for discovering a previously hidden structure in a set of unordered objects. In this case a true or natural grouping in the dataset is expected. However, the assignment of objects to the clusters and the description of these clusters are unknown. Arranging similar objects into clusters it is to consider unknown patterns with the aim that every cluster naturally represents a true type or category of objects.

In this thesis clustering methods are also used for the purpose of data reduction. With the use of Topic Maps (see chap. 4 et seq.) a simplified representation of the dataset is attempted, which allows for dealing with a manageable number of homogeneous groups instead of a vast number of single objects

[Kruse et al., 2007]. Only some mathematical criteria can decide on the composition of clusters when classifying datasets automatically. Therefore clustering methods are endowed with distance functions that measure the proximity of presented example cases. As a result a partition of the dataset into clusters regarding the selected proximities relation is yielded. Objects, which boast certain characteristics, are incorporated to the same cluster [Miyamoto et al., 2008]. For instance, consider the number of fingers and compare humans and monkeys. On this comparison the two species will be deemed to be similar. This kind of step leads to a monothetic, hierarchical classification, which is the basis of any hard clustering method [Govaert, 2009]. All the objects in the same cluster then feature a given number of attributes (e.g. all men are mortal). Therein partitioning clustering methods are different from hierarchical ones, the later arranging dataset in a nested sequence of clusters.

The clustering methods considered within this PhD thesis are basically partitioning methods: Given a positive integer, they aim at coming up with the most appropriate partition of the dataset X into c clusters based on proximity measures and they regard the space of feasible partitions into c clusters solely. Here, too, this could now and then be broadened. However, for the given situation, it is mentioned as an additional option.

3.2 Cluster Analysis for Object Data

This section showcases the whole process of clustering by analyzing the three major issues practical clustering has in it. As shown, clustering implies the partitioning of a dataset of objects into subsets so that objects in the same cluster are somehow similar [Govaert, 2009; Miyamoto et al., 2008]. Thereby the goal of the used methods is to gain a reduced representation of the initial dataset. The process of clustering mainly includes three issues: tendency assessment, clustering and its validation or evaluation [Bezdek et al., 2008].

Given a finite dataset $X \subseteq \mathbb{R}^n$, the task of clustering should be predefined as illustrated in figure 3.1. First the similarity or dissimilarity measurements should be designated. Measurements that quantify either the similarity or the dissimilarity between objects are generally referred to as proximity measurements. For that purpose section 3.2.1 introduces the most common proximity measurements. In accordance with the designated proximity measurement the dataset will be clustered into clusters such that objects with broadly matching characteristics are classified to the same clusters. Thereby the variables to which the clustering technique should be applied ought to be elected. According to the rules of the selected method the objects of the dataset are aggregated until eventually all objects are included in a cluster. In the end, in hard clustering each object belongs to exactly one cluster. However, in fuzzy clustering it is possible that an object can belong to numerous clusters (see sect. 3.3).

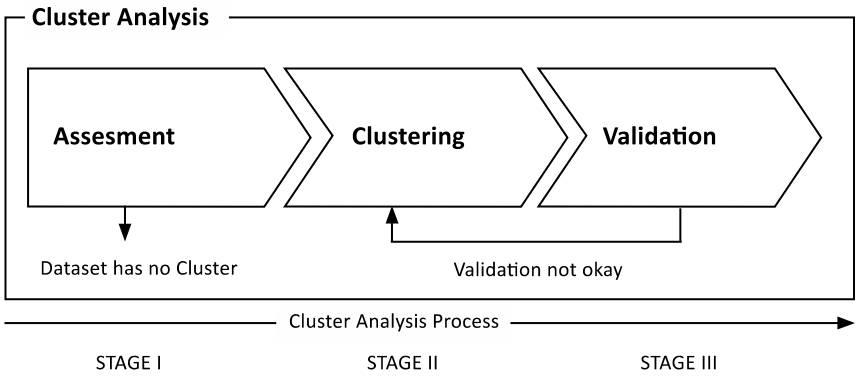

Figure 3.1: Three Problems of Cluster Analysis.

Following the process in figure 3.1, the next step is concerned with clustering the unlabeled dataset X. The expression learning used in the course of this thesis refers to perceiving good characteristics (and possibly prototypes) for the clusters in the dataset. For certain clustering algorithms, there is a parameter commonly referred to specifying the number of clusters to detect. Other algorithms do not require the specification of this parameter. Hence it must be defined which number of clusters is considered as the best in some way. Thereby the number of clusters and a clustering procedure must be selected.

Different practical approaches to define the optimal number of clusters are highlighted in section 3.2.2. Furthermore, once it is determined to look for clusters c, for example, then a model, whose measure of proximity may capture structure in the sense that a human may perceive it, should be selected to cluster the objects of the dataset into the chosen clusters. In doing that, the chosen clusters hinge on just the defined number of clusters and the selected proximity measurement. There are different clustering models and methods (i.e. algorithms). In this PhD thesis a distinction is made between a model, and methods used to solve or optimize it. The question what proximity measure to use is the centerpiece of all clustering models.

For didactical and simplicity reasons, section 3.2.3 introduces the reader to classical clustering; thereby a simple hard clustering model is presented. Since clustering is an unsupervised learning task, objects are not necessarily associated with characteristics that indicate its outcome. Thus no reference is provided to which the obtained results can be compared.

Different models and methods produce different partitions of the dataset, and it is not clear which one(s) may be most useful. So the next step is cluster validation. This refers to the review of the usefulness of generated clusters. To assess reliability and validity, various methods may be used: Therefore, in section 3.2.4, some of the most important methods are briefly introduced.

3.2.1 Determining the Proximity Measurement

Many clustering models and methods require the data to be represented as a set of proximities. Sometimes these proximities are the form in which the data naturally occur. In most clustering problems, however, each of the objects under investigation will be described by characteristics (e.g. variables or attributes). Hence, the first and possibly most important step in clustering is to define these proximities. Different kinds of definitions of proximity exist in literature [Backhaus et al., 2010; Baeza-Yates & Ribeiro-Neto, 1999]. These definitions are depending on the underlying types (e.g. continuous, binary, categorical or ordinal). Dependent on the level of measurement of the observed characteristics a variety of proximities have been developed. Figure 3.2 reveals some prominent examples of possible proximity measurements that are given a closer look in the following part.

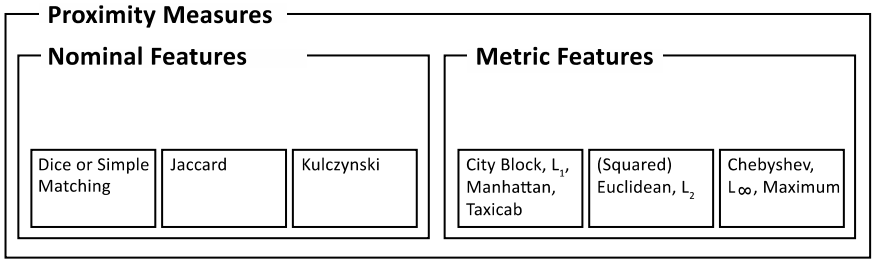

Figure 3.2: Selected Proximity Measurements.

Nominal features, which exhibit more than two characteristic values, are decomposed into binary (auxiliary) variables and to each characteristic either 1 (i.e. characteristic exists) or 0 (i.e. characteristic does not exist) is assigned. Thereby multi-categorical variables can be partitioned into binary variables and accordingly, in the following, similarity measurements for binary variables can be treated as a special case of nominal features. Note that large categories and categories of different size of such proximity measurements can lead to biases.

The notion of proximity, which is a quantitative measure of closeness, is a general term for similarity, dissimilarity and distance: Two objects are close when their dissimilarity or distance is small or their similarity large. More formally, dissimilarity on the set S can be defined as a function d from $S \times S$ to the real numbers such that:

1. $d(x, y) \geq 0$ for all x, y belonging to S
2. $d(\ x, x) = 0$ ifand only if $x = y$
3. $d(x, y) = d(y, x)$ for all x, y belonging to S

A dissimilarity satisfying the triangle inequality $d(x, z) \leq d(x, y) + d(y, z)$; $\forall x, y, z \in S$ is a distance. Note that the first condition is implied by

the others, since $2d(x,y) = d(x,y) + d(y,x) \geq d(x,x) = 0$, and that condition 1 and 2 together produce positive definiteness.

To compare the selected proximity measurements introduced in figure 3.2, the focus is on distance measurements. As shown, distance measurements help to identify the distance between two individual objects. The basis for distance measurement is a distance in $\mathbb{R}^n$: The L_p (or Minkowski) distance of order $p(\geq 1)$ between two points x and y, defined as:

$$d_{L_p}(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p}$$

Thereby $d_{L_p}(x,y)$ denotes the L_p distance of the objects $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$, and p the L_p constant, which is the critical factor in this equation to obtain. Note that p is an arbitrary number ≥ 1, but the L_p distance measurement is typically used with $p = 1$ or 2; the latter is the L_2 (or Euclidean) distance, while the former is known as L_1 (or City-block, Manhattan, resp. Taxicab) distance. In the limit for p tending to infinity:

$$\lim_{p\to\infty}\left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p}$$

the L_∞ (or Chebyshev, resp. Maximum) distance is obtained.

For a classification of data with qualitative characteristics (i.e. nominal data), coefficients such as for example the Dice (or Simple Matching), the Jaccard, and the Kulczynski coefficients, are widely used; these common nominal, but set-based ordinal distances are normally also applied in IR [Backhaus et al., 2010; Baeza-Yates et al., 2011].

Table 3.1: Definition of Distance Measurements.

Metric distance measurements	**Nominal distance measurements**
• L_1 *(or City-block, Manhattan or Taxicab (for p=1))* $d_{L_1}(x,y) = \sum_{i=1}^{n} \lvert x_i - y_i \rvert$	• *Dice (or Simple Matching)* $d_S(A,B) = 1 - \frac{2\lvert A \cap B \rvert}{\lvert A \rvert + \lvert B \rvert}$
• L_2 *(or (Squared) Euclidean (for p=2))* $d_{L_2}(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$	• *Jaccard* $d_J(A,B) = 1 - \frac{\lvert A \cap B \rvert}{\lvert A \cup B \rvert}$
• L_∞ *(or Chebyshev,* Maximum *(for p=∞))* $d_{L_\infty}(x,y) = \max_i(\lvert x_i - y_i \rvert)$	• *Kulczynski* $d_K(A,B) = 1 - \frac{1}{2}\left(\frac{\lvert A \cap B \rvert}{\lvert A \rvert} + \frac{\lvert A \cap B \rvert}{\lvert B \rvert}\right)$

The various measurements used in this thesis are illustrated in table 3.1; for the metric data the distances d are illustrated with points $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$ in $\mathbb{R}^n$. However, the three distances for nominal data are illustrated using sets A, B. Yet, there are numerous other measurements (often variations) not listed in this table. Many of these measurements can be found in [Backhaus et al., 2010].

Note that the L_p metric counts on dissimilarity measurements. As presented this implies that the bigger the measure is, the more dissimilar the single objects are. In many cases, not the distance but the similarity is desired: Commonly for nominal data the similarity is obtained by subtracting the particular distance from 1. Therefore in this case it is possible to calculate one from the other.

Now that a selected number of often used proximity measurements were presented, the next section introduces another frequent problem, the identification of a reasonable number of clusters in a given dataset.

3.2.2 Determining the Number of Clusters

Determining the number of clusters c in a dataset X is a problem in data clustering, and is a distinct issue from the process of actually solving the clustering problem itself. In this PhD thesis all used models are rooted in objective functions J, as mathematical criteria that quantify the fitness of clustering models that are made from the prototypes and the decomposition of its objects. The objective functions can be seen as cost functions that have to be minimized to obtain an ideal solution. Thus, for the following models the respective objective function's expressed criterion of optimality, the clustering task can be formulated as a function optimization problem. That is, its methods define the best decomposition of the dataset into the number of clusters by minimizing its objective function. The steps of the method follow from the optimization scheme that they apply to approach the optimum of J.

The idea of defining an objective function and have its minimization drive the clustering is rather cyclopedic. Aside from the basic algorithms various enhancements and alternations have been proposed with the goal to improve the clustering results with respect to particular problems (e.g. noise, outliers, etc.). Consequently, other objective functions have been tailored for these specific applications. However, regardless of the specific function that an algorithm is based on, the objective function is a measure of fitness [Kruse et al., 2007]. Thus it can be used to make comparisons between methods of a dataset that have been received by the same algorithm (holding the number of clusters fixed).

According to [Backhaus et al., 2010], it must be defined which number of clusters c is considered as most suitable. In general, there is no logically justifiable conception for a clustering and so it will often be tried to uncover the

data-inherent natural grouping. Thereby the determination of the optimal number of clusters c should be oriented towards statistical criteria and not on proper logic. Hence the correct choice of c is often ambiguous, with interpretations depending on the shape and scale of the distribution of objects in the dataset and the requested clustering. Apart from that, this choice of c can also affect performance of a clustering method. However, according to [Fu & Medico, 2007], none of the existing methods seems to perform significantly better than others when tested with different dataset.

Another issue is that increasing c without penalty will limit the number of errors in the resulting clusters, to the extreme case of zero if each object is considered its own cluster (i.e. when c equals the number of all considered objects $|X|$ of the dataset X). Intuitively, the optimal choice of c will strike a balance between maximum compression of the dataset using a single cluster, and maximum accuracy by assigning each object to its own cluster. If an analogous value of c is not detectable from past knowledge about the characteristics of the dataset, it must get picked in some way. Furthermore, if this task is relinquished to computers, which do not necessarily share the same views on object characteristics as humans do, a method to define the number of clusters must be specified. There are several methods for making this decision; the following section describes the most common ones briefly:

- *The simple rule method:* Following [Mardia et al., 1979], the simplest rule of thumb sets the number to $c \approx \sqrt{n/2}$, with n as the number of objects of the dataset X. Because of its simplicity of implementation, that rule of thumb is widely used in practice.

- *The elbow criterion method:* [Ketchen & Shook, 1996] look at the percentage of variance explained as a function of the number of clusters, whereby a cluster number c is picked with this method means that subjoining another cluster does not provide a considerably better modeling of the dataset. In doing so, the elbow criterion method is a plot of stress versus dimensionality, whereby the points in this plot usually form a convex pattern. The point at which an elbow or a sharp bend arises signifies an appropriate number of clusters. Increasing the number beyond this point is usually not worth the improvement.

- *The information criterion method:* These criterion methods normally measure the goodness of fit of an estimated model. The best-known information criterion methods are probably the Akaike or the Bayesian information criteria [Akaike, 1974; Schwarz, 1978]. These methods are used for model selection among a cluster of characteristic models with different numbers of characteristics. Thereby choosing a model to optimize an information criterion is a kind of regularization, involving the introduction of additional information in order to solve an ill-posed problem or prevent over-fitting for example. The information is usually in the form of penalty for complexity, such as restriction for smoothness or

bounds on the vectors space norm. These methods are coming from the concept of information entropy, offering a comparative measure of lost information using a certain method. These methods describe the tradeoff between bias and variance in the clustering construction, or loosely speaking between accuracy and complexity of the clustering [Miyamoto et al., 2008].

- *The jump criterion method:* The jump criterion method determines the number of clusters c that maximizes efficiency while minimizing error by information theoretic standards. This method arises from rate-distortion theory, a major branch of information theory. The strategy is to generate a distortion curve for the input data by running a standard clustering algorithm for all values of c between 1 and n, and then calculating a distortion curve of the resulting clustering. This distortion curve is then transformed by a negative power selected on the basis of the data dimensionality. Jumps in the resulting values indicate a reasonable selection for c, with the largest jump epitomizing the best selection [Sugar & James, 2003].

- *The silhouette criterion method:* The average shape of the dataset is another useful criterion for assessing the natural number of clusters. The silhouette criterion method provides a succinct graphical representation of how well each object lies within its cluster and how loosely it is matched to objects of the neighboring cluster, that is, the cluster whose average distance from the object is lowest. A silhouette close to 1 for example implies the object is in an appropriate cluster, whilst a silhouette close to -1 implies the object is in the wrong cluster [Rousseuw, 1987]. According to [Lleti et al., 2004] genetic optimization approaches that give rise to the largest silhouette are good variations.

- *The cross-validation method:* [Geisser, 1993; Ron, 1995] use the process of cross-validation (sometimes called rotation estimation) to analyze the clusters c. One round of cross-validation involves partitioning a dataset X into complementary clusters c, performing the analysis on one cluster (i.e. training set), and validating the analysis on the other cluster (i.e. test set). To reduce variability, multiple rounds of cross-validation are performed using different partitions, and the validation results are averaged over the rounds. In the end the cluster number is selected that minimizes the test set errors.

Now that the most common proximity measurements are known and several practical solutions to the assessment problem were introduced, the attention is applied to the main part of this chapter, the clustering itself.

3.2.3 CLUSTERING

With reference to [Kruse et al., 2007], in this section, a simple clustering model is envisaged as starting point for the later fuzzy extension of section

3.3. In this simple form of hard clustering, each cluster prototype is solely compromised of the center vectors, $C_i = (c_i)$, such that the objects assigned to a cluster are represented by a prototypical point in the object space. Following the relevant literature, as distances measure often d_{L_2} is used (see tab. 3.1); its description can be obtained from section 3.2.1. However, it is certainly possible that other distance measures can be used.

In classical clustering each object x_j in a given dataset X is assigned to exactly one cluster. Each cluster Γ_i is thus a subset of the given dataset, $\Gamma_i \subset X$. The whole set of clusters $\Gamma = \{\Gamma_1, \dots, \Gamma_c\}$ is required to be an exhaustive partition of the dataset X into c non-empty and pairwise disjoint subsets $\Gamma_i, 1 < c < n$. Such a partition is said to be ideal when the sum of the squared distance between the cluster centers c_i and the objects x_j assigned, is minimal [Kruse et al., 2007]. This is directly motivated from the requirement that clusters should be as homogenous as possible. Hence the objective functions for hard clustering—denoted by the subscript h—looks like:

$$J_h(X, U_h, C) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij} d_{ij}^2$$

where $C = \{C_1, \dots, C_c\}$ is the set of cluster prototypes, d_{ij} is the distance between x_j and cluster center c_i, U is a $c \times n$ binary matrix called partition matrix. The individual objects

$$u_{ij} \in \{0,1\}$$

indicate the binary allocation of an object to a cluster: $u_{ij} = 1$ if the object x_j is assigned to the prototype C_i, that is $x_j \in \Gamma_i$; and $u_{ij} = 0$ otherwise. To ensure that each object is allocated exactly to one cluster, it is required that:

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j \in \{1, \dots, n\}$$

This enforces full allocation and also avoids the trivial solution when minimizing J_h, which is that no object is assigned to any cluster: $u_{ij} = 0, \forall i, j$. Together with $u_{ij} \in \{0,1\}$ it is feasible that objects are assigned to one or more clusters while there are certain remaining clusters left empty. Since this is undesirable, it is usually required that:

$$\sum_{j=1}^{n} u_{ij} > 0, \forall i \in \{1, \dots, c\}.$$

J_h depends on the pairwise disjoint parameters that are the cluster centers c and the assignment of objects to clusters U [Kruse et al., 2007]. The clustering method minimizes J_h using an alternating optimization scheme.

Generally this scheme comes into operation when a function cannot be optimized straight, or when it is unpractical. The optimization parameters are separated into two (or even more) modes: Following one mode of characteristics (e.g. the partition matrix) is optimized holding the other mode(s) (e.g. the current cluster centers) fixed (and vice versa). This iterative updating scheme is then repeated. The main advantage of this method is that in each step the optimum can be calculated directly. By iterating the steps the joint optimum is approached, although it cannot be guaranteed that the global optimum will be reached [Kruse et al., 2007]. An algorithm may get stuck in a local minimum of the applied objective function J. However, alternating optimization is the commonly used parameter optimization method in clustering.

Here, at first the initial cluster centers are defined. This can be done randomly, in particular, by picking c random vectors that lie within the smallest (hyper-) cube that encloses the whole dataset; or by initializing cluster centers with randomly picked objects of the default dataset. Alternatively, more sophisticated initializing methods can be used as well, for example Latin hypercube sampling [McKay et al., 1979]. Then the parameters C are held fixed and clustering assignments U are ascertained that minimize the quantity of J_h. In this step each object is assigned to its nearest cluster center:

$$u_{ij} = \begin{cases} 1, if\ d_{ij} = min_{l=1}^{c} d_{lj} \\ 0, \quad otherwise \end{cases}$$

Here it is assumed that for every j one of the numbers d_{lj} is strictly smaller than all the others, so that the value 1 is only assigned to one u_{lj}. Any other allocation of objects than to its closest cluster would not minimize J_c for fixed clusters. Then the data partition U is held fixed and new cluster centers are calculated as the mean of all data vectors assigned to them, since the mean minimizes the sum of the square distance in J_h. The computation of the mean for each cluster is stated more formally:

$$c_i = \frac{\sum_{j=1}^{n} u_{ij} x_j}{\sum_{j=1}^{n} u_{ij}}$$

The two steps are iterated until no change in C or U can be observed. Then the algorithm terminates, yielding final cluster centers and data partitions. Now that the assessment of objects of the dataset and the clustering of which have been introduced, the established cluster should be validated somehow. Therefore the next section presents three different criteria for validation or evaluation of a created clustering.

3.2.4 VALIDATION OF THE CLUSTERS

An important topic related to clustering is that of cluster validation or evaluation, that means the assessment of the obtained cluster quality. Major cluster validity approaches include the evaluation of the tradeoff between cluster compactness, separability and stability-based approaches. Such approaches can be used to compare how well a clustering method performs on a dataset. These measures are usually associated to the type of criterion being considered in assessing the quality of a clustering method.

- *Internal criterion:* Clustering evaluation methods that abide by internal criterion assign the best score to the method that produces clusters with high similarity within a cluster and low similarity between clusters. According to [Manning et al., 2008], a drawback of using internal criterion in cluster evaluation is that high scores on an internal measure do not necessarily result in effective IR applications.

- *External criterion:* Clustering evaluation methods that abide by external criterion compare the clustering results against some external benchmark. Such benchmarks are comprised of a set of pre-classified items, and these sets are often created by experts. These types of evaluation methods measure how close the clustering is to the predetermined benchmark clusters. However, recent debates were held whether this is adequate for real, or only synthetic datasets with a factual ground truth [Färber et al., 2010]. Since clusters can include an internal structure, the attributes present may not allow separation of clusters or the classes may contain anomalies. Besides, external criteria do not harmonize very well with unsupervised testing. Yet, for completeness they are revealed here.

- *Relative criterion:* Cluster evaluation methods that integrate relative criterion assess the clustering method in terms of user need. For example, a clustering algorithm may, based on different internal and external criteria, perform excellent, but the algorithm may be unnecessarily slow. If the user seeks a quick clustering response, a faster algorithm that performs slightly poorer on the internal criterion may be preferred. This evaluation criterion is straighter and requires the definition of the user need.

For the relative criterion there exists no consistent evaluation method. However, for internal and external evaluation methods several validation criteria were developed. As example, table 3.2 illustrates two indices in each case to assess the quality of clustering algorithms, based on internal and external criteria. More complete lists can be found by [Backhaus et al., 2010; Bezdek et al., 2008; de Oliveira & Pedrycz, 2007].

Note that in the examples of internal criteria, for a good clustering the Davies–Bouldin index I_{DB} should be as low as possible, whereas the Dunn index I_D should be as high as possible. Since methods that produce clusters with low intra-cluster distances (i.e. high intra-cluster similarity) and high inter-cluster distances (i.e. low inter-cluster similarity) will have a low I_{DB}, the clustering method that produces a collection of clusters with the smallest I_{DB} is considered the best based on this criteria. Since internal criterion seek clusters with high intra-cluster similarity and low inter-cluster similarity, by contrast, methods that produce clusters with high I_D are more desirable.

Table 3.2: Example Criteria for Clustering Evaluation.

Examples of Internal Criteria	
Davies–Bouldin Index $I_{DB} = \frac{1}{n}\sum_{i=1}^{n} \max_{i \neq j}\left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)}\right)$	Where: n is the number of clusters c_i is the centroid of cluster i σ_i is the average distance of objects in i to c_i $d(c_i, c_j)$ is the distance between centroids c_i and c_j
Dunn Index $I_D = \min_{1 \leq i \leq n}\left[\min_{1 \leq j \leq n, i \neq j}\left(\frac{d(i,j)}{max_{1 \leq k \leq n} d'(k)}\right)\right]$	Where: $d(i,j)$ is the distance between clusters i and j $d'(k)$ is the diameter of cluster k
Examples of External Criteria	
Rand Index $I_R = \frac{TP + TN}{TP + FP + FN + TN}$	Where: TP is the number of true positives TN is the number of true negatives FP is the number of false positives FN is the number of false negatives
F-Measure $F_\beta = \frac{(\beta^2 + 1) \cdot P \cdot R}{\beta^2 \cdot P + R}$	Where: $\beta(\geq 0)$ is the recall parameter $P(= TP/TP + FP)$ is the precision rate $R(= TN/FN + TN)$ is the recall rate Note: when $\beta = 0, F_0 = P$ when $\beta \to \infty, F_\beta \to R$

The True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) used for the Rand index I_R and F-measure originate from statistical hypothesis testing. Therein, type I and type II errors refer to incorrect conclusions that can be drawn during a test. In any test there are

four basic possibilities: Two that are correct — something is true and a test says it is true (TP); something is false and a test says it is false (TN) — and two possibilities which are errors — something is false but a test says it is true (FP); and something is true but a test says it is false (FN). Here it is tested if two objects of the dataset X are in the same cluster: For instance, this implies for TP that a pair of objects of the dataset is classified to the same cluster by both, the tested and the reference clustering.

In the F-measure, Precision (P) can be seen as a measure of exactness, whereas Recall (R) is a measure of completeness. When using P and R, the set of possible labels for a given instance is divided into two subsets, one of which is considered relevant for the purposes of the metric. R is then computed as the fraction of correct instances among all instances that actually belong to the relevant subset, while P is the fraction of correct instances among those that the algorithm believes to belong to the relevant subset [Backhaus et al., 2010; Baeza-Yates & Ribeiro-Neto, 2011].

3.3 Introduction to Fuzziness

As illustrated, in their original forms clustering methods look for a predetermined number of clusters in a dataset where the center vector features each of the clusters. Thereby each object is assigned to precisely one cluster, generating exhaustive partitions of the dataset into non-empty and pairwise-disjoint subsets. This results from traditional set theory forcing objects to be either in a set or not. Fuzzy set theory relaxes the requirements of traditional set theory. In fuzzy set theory objects can belong to more than one class and even with different degrees of membership to the different classes. A short primer to fuzzy set theory is therefore presented in section 3.3.1 as an introduction to understand (more easily) subsequent fuzzy clustering.

After section 3.3.1, the gained insights will be applied to clustering and thereby enhance hard clustering to fuzzy clustering. Since hard assignment of objects can be inadequate in the presence of objects that are almost equally distant from two or more clusters, the mentioned objects can represent hybrid-type or mixed objects, which are (more or less) equally similar to two or more types. A hard partition arbitrarily forces the full assignment of objects to one of the clusters, although they should (almost) equally be assigned to all of them. The fuzzy clustering methods presented in section 3.3.2 relax the requirement that different objects of a dataset have to belong to only one cluster. The shift from hard to gradual assignment of objects to clusters for the purpose of more expressive data partitions founded the field of fuzzy clustering. So gradual cluster assignments can reflect the on hand cluster structure in a more natural way, particularly when clusters overlap. Then the memberships of objects at the overlapping boundaries can express the ambiguity of the cluster assignment.

3.3.1 FUZZY SET THEORY

This introductory section aims to present the main concepts and mathematical notions of the fuzzy set theory (also referred to as fuzzy logic or fuzzy logic theory), which are necessary for the understanding of the thesis. Proposed by [Zadeh, 1965], the fuzzy set theory is an extension of classical set theory based on intuitive reasoning, that takes into account vagueness, imprecision and uncertainty. The boundaries of classical sets are required to be drawn precisely and, therefore, set membership is determined with complete certainty. An object is either definitively a member of a set or definitively not. This hard distinction is also reflected in classical logic, where each proposition is treated as either definitively true or false. Tertium non datur. This logic results from Aristotle's stated law of excluded middle, that for any proposition Π specifies, that either the proposition is true, or its negation is; $\Pi \vee \neg\Pi$, where Π is a model for any proposition such as penguins live in southern hemisphere, an apple is worm-eaten, and so on.

Exemplified, this logic is applied to set theory and the set of alphabet vowels is defined. Logically, the set of consonants exclude the set of vowels, because per definition, a letter is either a consonant or a vowel. However, in English, the letter y is sometimes a vowel and sometimes a consonant. For example, in the word my, y is a vowel, but in the word yours, it is not. The question is now, whether y belongs in the vowels set, or if it belongs in the consonants set. The answer is unclear because y does not fit seamlessly into either set, but rather in both [Smithson & Verkuilen, 2006]. This means, of course, that the rule separating vowels and consonants does not lead to a mutually exclusive classification of letters as proposed by the dichotomy between vowels and consonants. Ergo the letter y violates Aristotle's law of excluded middle that is assumed when $\Pi \vee \neg\Pi$ is defined.

It is awkward to think sharp about even this simple illustration, but the hitch is similar to vagueness those faced in the case of compiling datasets and making inference about objects in these datasets. Therefore classical set theory is often not enough for handling vagueness in the rule that assigns objects to sets. Mathematical objects generally can be defined precisely; real objects by contrast cannot be defined so easily.

Fuzzy sets are designed to handle a particular kind of vagueness, which results when a characteristic is established that holds by objects to varying degrees. Vagueness is easiest to see with reference to a classical paradox, the Sorites: Imagine, for example, a heap of sand. If one grain of sand is removed from this heap, the residual pile is still a heap. Arguing by a conceivably fallacious appeal to mathematical induction, therefore another grain of sand is removed with the result, that there is still a heap. This is iteratively repeated and, eventually, however, there is so little sand that no one would be willing to call whatever is left a heap. Thus, the definition of heap is not precise. It is subject to vagueness because nowhere in the action there is a point that

separates things into two clear distinguishable states: Heap and not-heap [Smithson & Verkuilen, 2006]. Furthermore, consider as another example again the marriage of Heather Pegg and Jony Ive (see chap. 2). Their marriage — and all other marriages certainly as well — videlicet brings with it another paradox: I cannot live with her, and I cannot live without her. Both statements are (to a certain degree) true. The dynamic between those statements is (along with other things) what keeps marriage interesting. That is exactly what fuzzy sets deal with. Down with both statements (i.e. with and without) in fuzzy set theory certain membership degrees come along.

A fuzzy set is grounded on a classical set, but it adds a key element to it: A numerical degree of membership of an object in the set, ranging from 0 to 1. Formally, a fuzzy set is built from a reference set called universe of discourse Ω. This reference set is never fuzzy. A fuzzy set A (in Ω) is the graph of a function $\mu_A: \Omega \rightarrow [0,1]$, called the membership function of A. The value $\mu_A(x)$ for x in A stands for the degree of membership of x in A. Note that in this thesis the function is identified with the graph and that a domain may refer to Ω, but it also can be defined in the sense of some mathematical region such as the real line or an interval representing the array of spectrum.

The membership function is an index of sethood that quantifies the degree to which an object x is a member of a certain set. Unlike probability theory, degrees of membership do not inevitably have to add up to 1 across all objects. As a result many or few objects in the set could have high membership. However, an object's membership in a set and the set's algebraic complement must still sum to 1 [Smithson & Verkuilen, 2006]. The main difference between classical and fuzzy set theory is that the latter allows to partial set membership. A classical hard set, then, is a fuzzy set that constrains its membership values to $\{0,1\}$, the termini of the unit interval. Fuzzy set theory projects vague phenomena by assigning any object a weight given by the value of the membership function, measuring the extent to which a rule that an object belongs to set A is assessed to be true.

At this, it should not be left unmentioned, that fuzzy set theory includes traditional as well as additional set operation scope. Accordingly, like classical set theory, fuzzy set theory includes operations union, intersection, complement, and inclusion, but also operations that have no classical counterpart, such as the modifiers concentration and dilation, and the connective fuzzy aggregation. Since these operations are not needed in the PhD project, these concepts are not introduced at this point. Anyway, for interested readers, [Klir et al., 1997; Smithson & Verkuilen, 2006; Zimmermann, 2001] are highly recommended. In the following section the ideas of fuzzy set theory will now be applied to clustering.

3.3.2 FUZZY CLUSTERING

Hard and fuzzy clustering algorithms differ in how they assign data to a cluster, therefore in what type of data partitions they form. The fuzzy set theory applied to clustering results in fuzzy clustering that allows gradual memberships of objects to clusters measured as degrees in [0,1]. This gives the flexibility to express that objects can belong to more than one cluster. Furthermore, the membership degrees provide much finer-grained values of detail of a model. Aside from assigning an object to clusters in shares, membership degrees can also reveal how ambiguously or definitely objects should belong to a certain cluster. For example, figure 3.3 illustrates homonymy between Oppie five II (i.e. Apple's Chief Financial Officer (CFO) Peter Oppenheimer's yacht) bow and the archery weapon that uses elasticity to shoot arrows into apples. By implication the term `:Bow` belongs to both, the cluster `:Weapon` and the cluster of `:Ships` (i.e. with a membership degree of 0.6).

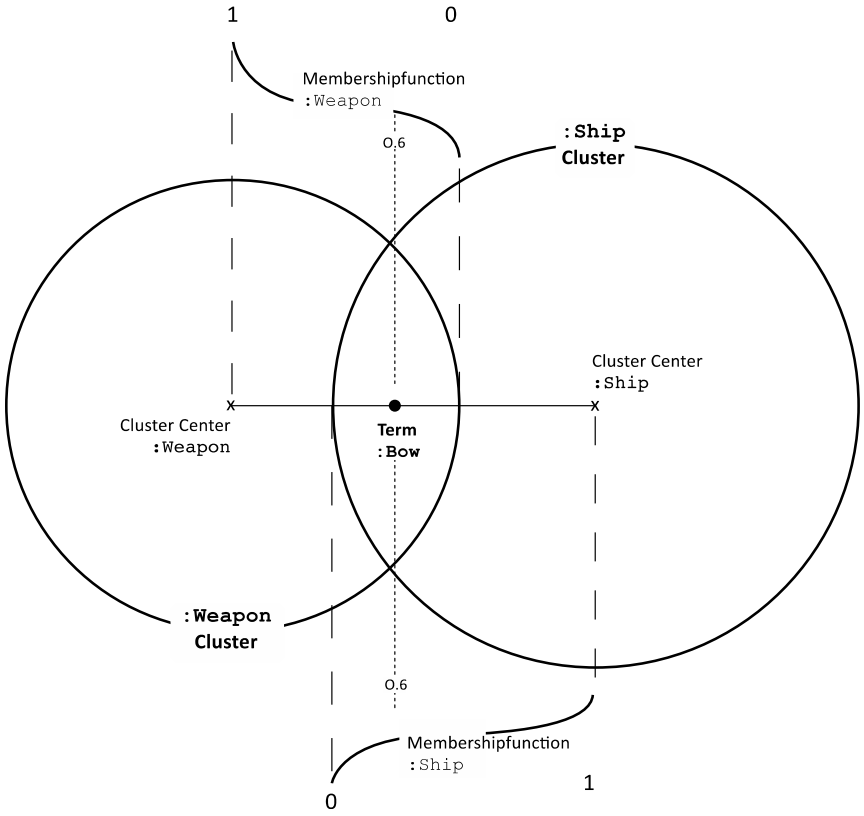

Figure 3.3: Bow Membership Degree Example.

The concept of membership degrees is substantiated by the definition and interpretation of fuzzy sets. Thus, fuzzy clustering allows fine grained denouncements in the form of fuzzy partitions of the set of given examples X. The clusters Γ_i of data partitions have been classical subsets so far; following

suit [Kruse et al., 2007] they are represented by fuzzy sets $\mu_{\Gamma_i}$of the dataset X in the following. Complying with fuzzy set theory, the cluster assignment u_{ij} is now the membership degree of an object x_j to cluster Γ_i, such that: $u_{ij} = \mu_{\Gamma_i}(x_j) \in [0,1]$. Since memberships to clusters are fuzzy, there is not a single label that is indicative to which objects belongs. Instead, fuzzy clustering methods dedicate a fuzzy characteristic to each object x_j that reveals its memberships to the c clusters:

$$u_j = (u_{1j}, \dots, u_{cj})^T$$

The $c \times n$ matrix $U = (u_{ij}) = (u_1, \dots, u_n)$ is then called a fuzzy partition matrix. This allows, based on the fuzzy set notion, a better-suited handling of ambiguity of cluster assignments when clusters are poorly outlined or overlapping. Up to now, the general definition of fuzzy partition matrices leaves open how assignments of objects to more than one cluster should be uttered with membership degrees. Furthermore, after [Kruse et al., 2007; Miyamoto et al., 2008], the degrees of belonging to a certain cluster are still ambiguous, that means, the solution space (i.e. set of allowed fuzzy partitions) for fuzzy clustering is not yet specified.

Let X be the set of given objects and let c be the number of cluster $(1 < c < n)$ represented by the fuzzy sets $\mu_{\Gamma_i}, (i = 1, \dots, c)$. Then $U_f = (u_{ij}) = \left(\mu_{\Gamma_i}(x_j)\right)$ is called a fuzzy partition—labeled with the subscript f for fuzzy—of X if the conditions:

$$\sum_{j=1}^{n} u_{ij} > 0, \forall i \in \{1, \dots, c\}, and$$

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j \in \{1, \dots, n\}$$

hold. The $u_{ij} \in [0,1]$ are constructed as the membership degree of x_j to cluster Γ_i relative to all other clusters.

The first condition ensures that no cluster is empty. This corresponds to the requirement in classical clustering that no cluster, represented as (classical) subset of X, is empty. The second condition guarantees that the sum of the membership degrees for each object equals 1. This means that each object receives the same weight with respect to all other objects, and, therefore, that all objects are (equally) included into the cluster partition. This is related to the constraints in classical clustering that partitions are formed exhaustively. As a consequence of both conditions it is possible that no cluster contains the full memberships per object [Bezdek et al., 2008; Miyamoto et

al., 2008]. Thus the membership degrees for a given object formally correspond to the probabilities of its being a member of the respective cluster.

After this specification of probabilistic partitions, an objective function for the fuzzy clustering task can be assigned. Certainly, the closer an object is located to a cluster center, the higher its degree of membership to this cluster should be. Following this rationale, it is possible to say that the distances between the cluster centers and the objects (strongly) assigned to it should be minimal. Hence the problem to divide a given dataset into *c* clusters can (again) be stated as the task to minimize the squared distances of the objects to their cluster centers, since, of course, the maximization of membership degrees is intended [Bezdek et al., 2008; Kruse et al., 2007; Miyamoto et al., 2008]. The fuzzy objective function J_f is thus based on the least sum of squared distances just as J_h of the hard clustering. More formally, a fuzzy cluster method of a given dataset X into c cluster by taking into account the weighting exponent m(>1), is defined to be optimal when it minimizes the objective function:

$$J_f(X, U_f, C) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d_{ij}^2 ,$$

under the above mentioned two conditions that have to be satisfied for probabilistic membership degrees in U_f. The first constraint averts the trivial solution of minimization problem, $u_{ij} = 0, \forall i, j$. The normalization condition leads to a distribution of the weight of each object over the different clusters. Since all objects have the same fixed amount of membership to split between clusters, the normalization constraint implements the known partitioning property of any probabilistic fuzzy clustering method. In fuzzy literature the weighting exponent m is called fuzzifier. The exponentiation of the memberships with m in J_f can be seen as a function g of the membership degrees $g(u_{ij}) = u_{ij}^m$, that results in a generalization of the least squared error function. The actual value of m affects the fuzziness of the classification. In [Dunn, 1974] it has been shown for the case $m = 1$ (that is when J_c and J_f become identical) that cluster assignments stay hard when minimizing the target function: even though they are allowed to be fuzzy, that means they are not limited to {0,1}. To come by the requested fuzzification of the resulting probabilistic partition, the function $g(u_{ij}) = u_{ij}^2$ has been suggested by [Dunn, 1974]. The generalization for exponents $m > 1$ that lead to fuzzy memberships has been proposed by [Bezdek, 1973]. With higher values for m, the boundaries between clusters become softer, with lower values they get harder. According to [Kruse et al., 2007], usually $m = 2$ is chosen. Aside from the standard weighting of memberships with u_{ij}^m other functions g that can go for fuzzifiers have been investigated [Kruse et al., 2007].

The objective function J_f is alternately optimized, first the membership degrees are optimized for fixed cluster parameters, and second the cluster prototypes are optimized for fixed membership degrees:

$$U_\tau = j_U(C_{\tau-1}), \tau > 0 \; and$$

$$C_\tau = j_C(U_\tau).$$

In each of the two steps the optimum can be calculated simply by applying the parameter update equation j_U and j_C for the membership degrees and the cluster centers, respectively [Kruse et al., 2007]. The updated formulae are derived by roughly putting the derivative of the objective function J_f with reference to the parameters to optimize equal to zero (taking into account the second condition). The resulting equations for the two iterative steps form the fuzzy clustering algorithm.

According to [Kruse et al., 2007], the membership degrees have to be chosen according to the ensuing update formula that is unmatched to the culled distance measure:

$$u_{ij} = \frac{1}{\sum_{l=1}^{c} \left(\frac{d_{ij}^2}{d_{lj}^2} \right)^{\frac{1}{m-1}}} = \frac{d_{ij}^{-\frac{2}{m-1}}}{\sum_{l=1}^{c} d_{lj}^{-\frac{2}{m-1}}}.$$

In case there exists a cluster i with no distance to an object x_j, then $u_{ij} = 1$ and $u_{lj} = 0$ for all other clusters $l \neq i$. The above equation illustrates the comparative characteristic of the probabilistic membership degree. It depends not only on the distance of the object x_j to cluster i, but also on the distance between this object and other clusters.

The update formulae j_C for the cluster parameters depend, on the parameters used to describe a cluster (e.g. location, shape, size) and on the chosen distance measure (i.e. relationship). Therefore a general update formula cannot be given. In the case of the basic fuzzy clustering model the cluster center vectors serve as prototypes, while an inner product norm induced metric is applied as distance measurement. Consequently the derivations of J_f with reference to the center yield:

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m}$$

The choice of the optimal cluster center points for fixed memberships of the dataset to the clusters has the form of a generalized mean value computation.

The general form of the alternating optimization scheme of linked equations opens with an update of the membership matrix in the initial iteration of the algorithm ($\tau = 1$). The first calculation of memberships is based on a first set of prototypes C_0. Even though the optimization of an objective function could mathematically also start with an initial but valid membership matrix (i.e. fulfilling the two constraints), a C_0 initialization is easier and therefore common practice in all fuzzy clustering methods [Kruse et al., 2007].

3.4 Applying Fuzzy Clustering to the Social Semantic Web

In the Semantic Web RDFa is premised on metadata. Thereby HTML5, for example, allows creators including metadata in HTML into their website directly unlike the Social Web way of annotating semantics by folksonomies [Breslin et al., 2009; Pilgrim, 2010]. In social bookmarking (also called collaborative tagging), users assign tags to resources shared with other users, which gives rise to a type of information organization that emerges from this crowdsourcing process. The resulting information structure can be seen as reflecting the collective knowledge (or collective intelligence) of a community of users. Actually, its ease of use pushes folksonomies broadly (see chap. 2). A large number of annotations through tags improve annotated Web data quality. Like teaching a child, using the law of large numbers can stabilize concepts in Social Web. For example, if numerous people annotate the same source with the same tags the relationship grows stronger. This is comparable to train a child a concept and thereby programming its brains neuronal pattern in relation to the concept [Spitzer, 2000]. In the course of this, the law of large numbers is critical because it guarantees stable long-term results for random events [Breslin et al., 2009; Pilgrim, 2010]. Already tags used together can produce an emerging relationship between them.

Tags are generally chosen informally and personally by a creator or by viewers (depending on the system used to describe the item) to aid searching. For this reason, tags are simple to create but generally lack a formal grounding, as intended by the Semantic Web [Voss, 2007]. Through tags, value is added by structuring the information and ranking it in order of relevance to ease query searches, as outlined by [Orio, 2010]. Thereby the tags must not necessarily be assigned by humans. Besides their own tag inference technique, [Budura et al., 2009] presents different ongoing work of automatic tag propagation. In addition, also facial recognition software (e.g. Google Find my Face) can support this propagation. In this PhD project, folksonomies are used as a starting point to harvest collective knowledge, which is then linguistically normalized and converted into a computer-understandable ontology.

As presented in chapter 2, this can provide ontologies a common terminology, which can be used to model a domain. After [Breslin et al., 2009] they are characterized by two features: First, in formal languages they are expressed with semantic meaning and second they represent a shared understanding

of a domain within a community (e.g. of experts or public groups). A domain comprises the types of objects and concepts that exist and their properties and relations.

Through repetitive harvesting of tags from folksonomies (e.g. through a daily, hourly or even real-time revisit of identified folksonomies through Web agents; see chap. 6), a tagspace (a set of associated tags with related weights) can be created in which semantic closeness is represented by distance d. To achieve an allied tagspace (where all harvested tags are related to each other), it is essential to establish tags and their relationships to each other [Kaser & Lemire, 2007; Budura et al., 2009]. These tags and their relationships are calculated using the introduced proximity measurements. The easiest way to find the similarity between two tags is to count the number of co-occurrences that is the number of times the two tags are allocated to the same source [Hassan-Montero & Herrero-Solana, 2006]. However, there are other measurements to establish similarity, such as locality-sensitive hashing (where the tags are hashed in such a way that similar tags are mapped to the same set with a high probability) and collaborative filtering (where several users define tags and their relations jointly). Each of these methods produces relationships among tags, and each offers a semantically consistent picture in which the tags are related to each other to some degree [Suzuki & Setsuo, 2004].

At present, the intention for the semantic Social Web (i.e. the grassroots-driven Social Semantic Web) is to adjust the bottom-up endeavor of the Social Web in a top-down fashion [Cardoso, 2007]. The fundamental aim is a stronger knowledge representation, as can be achieved with folksonomies, for example. Fuzziness can overcome the gap between folksonomies and ontologies because fuzziness corresponds to the natural way in which humans think [Meier et al., 2008] and it is, thus, suitable for characterizing vague information and helps to more efficiently handle real-world complexities. One possible way to use these advantages is through fuzzy clustering, which allow modeling of the uncertainty associated with vagueness and imprecision through mathematical models. As presented, fuzzy clustering generally deals with imprecision, uncertainty, partial truth and approximation.

To build the ontology, the tagspace will be clustered first with random initialization by a fuzzy clustering algorithm into c pre-computed classes [Portmann & Meier, 2010]. A fuzzy clustering algorithm classifies all collected tags of the tagspace. As presented, assigning cluster numbers ex ante is a common problem in clustering. The previously presented methods for defining the optimal cluster number can produce relief. In fuzzy clustering, each point has a degree of belonging to a class using fuzzy logic rather than belonging to one particular class. Thus, points on the edge of a class may participate to a less significant degree than points in the center of a class. The degree of membership is in the interval between 0 and 1. The greater this

membership is, the stronger the membership of an element (i.e. tags) to the class will be.

The relationship (along with the distance) to the other classes and also to the tags of each class is accomplished using RDFS and OWL. Thus, the number of classes can be determined by various methods (see sect. 3.2.2). Because this is fuzzy clustering, it is possible for each tag to belong to one or more classes with different degrees of membership. Accordingly, it is also possible that linguistic issues such as synonyms and homonyms can be identified and fuzzily resolved in a manner of speaking. Synonyms (from Greek meaning metonymic with) are different words with almost identical or similar meanings, compared to homonyms (from Greek meaning having the same name) that are tantamount to a group of words, which share the same spelling and the same pronunciation but have different meanings. Besides, following [Saeed, 2008] homonyms can be partitioned simultaneously to homographs (i.e. words that share the same spelling, irrespective of their pronunciation) and to homophones (i.e. words that share the same pronunciation, irrespective of their spelling).

Long story short, because the harvested synonymy tags `:Bough` and `:Branch` are related to each other (i.e. shows high similarity), by approximation they can be identified to belong to the class `:Tree`. In addition, as every tag can belong to different classes, it is possible that the homonymy tag `:Bow` can belong to either the class `:Ship` or the class `:Weapon`. Because the harvested tags are normalized, it is furthermore possible to spot homophones such as `:Bow` and `:Bough`. Figure 3.3 illustrates these small linguistic examples as ontology.

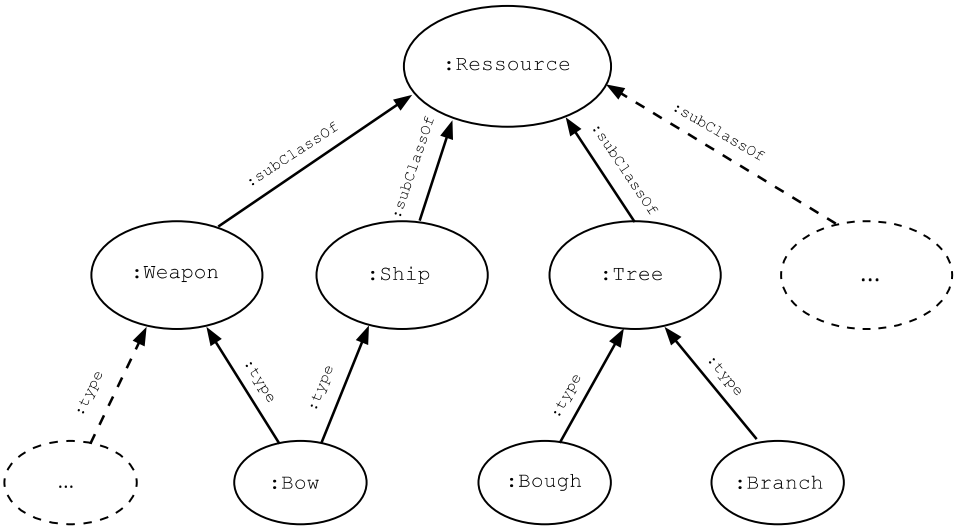

Figure 3.4: Fragment of an Ontology.

The creation of a fuzzy grassroots ontology—named after the social movement driven from bottom up—is a process where generated classes are stored in a (graph) database. This process is iteratively repeated. During these iterations, all of the generated classes of the fuzzy grassroots ontology—with the center naming it—are collected and stored in the database by an ontology knowledge administration system (e.g. AllegroGraph, Jena or KAON). More about the creation of the fuzzy grassroots ontology can be found in chapter 6 and 7. However, using this established ontology, it is possible for both humans and computers to recognize dependencies. For example, by trailing up a `:Watercraft` ontology, it is feasible to deduce that `:Boat`s are related to `:Ship`s. Furthermore, it is possible to recognize that, besides `:Watercraft`s, there are also `:Aircraft`s, for example [Portmann et al., 2012].

Compiling ontologies in the Social Semantic Web follow a power law distribution, whereby the social Semantic Web makes up a few precise ontologies (i.e. hard) that (maybe) cannot be sufficiently generalized for everyday applications and the semantic Social Web constitutes the long tail of vague (i.e. fuzzy) ontologies that are (possibly) not precise enough for specific applications. Data provided in the latter (on which this PhD project focuses) are typically based on vague human perceptions such as time (e.g. today, yesterday, last seven days, etc.), tags (e.g. related terms, topics, etc.), locations (e.g. America, Asia, Europe, etc.), groups (e.g. friends, unknowns, etc.), and other characteristic of physical and mental objects.

Now, to implement a Social Semantic Web, the computers must be given the ability to recognize and understand perceptions. To this end, [Zadeh, 2002]'s computational theory of perceptions take human vagueness into account. Hence, in the course of this, fuzziness plays a key role since it is especially well suited for handling spontaneously arising grassroots structures (see sect. 3.3). In contrast to a standardized ontology, the on this insight based fuzzy grassroots ontology is thus useful for wielding natural bottom-up structures and therefore to add semantics to Social Web data.

Applications for fuzzy grassroots ontology in the Social Semantic Web facilitate information sharing, enable more sophisticated search engines, support intelligent agents and the pushing of data, minimize data loss or repetition, and help with the discovery of resources by enabling field-based searches. During the PhD project, the fuzzy grassroots ontology was—beside the FORA framework—instituted in three different other projects:

- *eGlossary:* A proposal of an innovative electronic glossary (eGlossary) project for the dpunkt Heidelberg publishing company. This glossary presents explanations syntonic to a users' state of knowledge. However, in a general sense, a glossary contains explanations of concepts relevant to a certain field of study or action. In this sense, the term is related to the notion of ontology. The eGlossary builds on the fuzzy grassroots on-

tology to provide the user additional concepts related to the term searched after. This can be thought of as an automated form of Wikipedia's see also concept [Martinez, 2010; Martinez, 2011].

- *InRiNa*[25]: An Innovation- and Risk-Navigator project compiled together with the Innovation Society Ltd. St. Gall. In this project the fuzzy grassroots ontology provides related Nano-technological components according a query. The creation of the fuzzy grassroots ontology thereby rests not on the Social Web but on documents describing Nano-technology. Through building inverted indices of these underlying documents a tag-space to create the fuzzy grassroots ontology is generated. An inverted index is an index data structure storing a mapping from content (e.g. words or numbers), to its locations in a database file (or in one or a set of documents). The purpose of the inverted index is to allow fast full-text searches, at a cost of increased processing when a document is added to the database. This enables specialists to preserve a more consistent picture of a domain [Wehrle & Portmann, 2012].

- *Prometheus*: A building intelligence project initialized together with the iHomeLab[26]. Prometheus introduces a vision for further enhanced Web of Things (WOT) services (where everyday devices and objects are connected by fully integrating them to the Web). Based on a variety of data (e.g. location data, indoor and outdoor conditions, as well as fuzzy grassroots ontology-backed search queries) the Prometheus framework is intended to support users with helpful recommendations and information preceding a search for context-aware data. Adapted from artificial intelligence concepts, Prometheus proposes user-readjusted answers on umpteen conditions [Portmann et al., 2010; Andrushevich et al., 2011].

3.5 Further Readings

An introduction to various methods of multivariate data analysis can be found in [Backhaus et al., 2010] as well as in [Govaert, 2009]. Thereby the former roots are in marketing, whereby the latter is coming from digital and image processing techniques. [Bezdek et al., 2008; Miyamoto et al., 2008; de Oliveira & Pedrycz, 2007] introduce fuzzy clustering for numerous applications such as pattern recognition, image processing and data mining.

A detailed sketch of proximity measurements in general can be found by [Backhaus et al., 2010] and for measurements used in IR the book of [Baeza-Yates & Ribeiro-Neto, 1999] is a good source. [Backhaus et al., 2010] help finding the number of clusters and [Bezdek et al., 2008; Govaert, 2009] describe different approaches how to validate found (fuzzy) clusters.

For an introduction into fuzzy logic the following books can be recommended: An introduction to many-valued and fuzzy logic by [Bergmann, 2008] and a first course in fuzzy logic by [Nguyen & Walker, 2005]. Fuzzy set theory can be extracted from the following recommended sources: [Klir et al.,

1997; Smithson & Verkuilen, 2006; Zimmermann, 2001]. Thereby these books also cover introductions in set theory, vagueness, as well as human kinds and fuzziness.

The conversion of metadata to fuzzy grassroots ontology can be found by [Portmann et al., 2012; Portmann, 2011a; Portmann & Meier, 2010; Portmann & Kuhn, 2010]. A sketch of the eGlossary project [Martinez, 2010] and [Martinez, 2011]. The InRiNa project is detailed in [Wehrle & Portmann, 2012]. Last but not least, the Prometheus framework is covered in [Andrushevich et al., 2011] and [Portmann et al., 2010].

Part II

Field of Application

4 Online Reputation Analysis

"It takes many good deeds to build a good reputation, and only one bad one to lose it."

—Benjamin Franklin

To buy a good reputation is virtually impossible. To the contrary, reputation is earned over time, based on character, words, and actions. It is hard to build, easy to lose and it matters a lot. Because of reputation, organizations and people fail or succeed and brands live and die on it. Sticks and stones may break bones but a bad reputation can put an organization out of business. Unfortunately, so far, there is no multidisciplinary accepted definition of the term [Barnett et al., 2006; Eberl, 2006; Eisenegger & Imhof, 2009; Walsh, 2006]. The main reason for this is that reputation is a social construct that is formed by people: so everyone has an inkling of it. Likewise reputation is studied in umpteen disciplines with different approaches. Despite the variation of definitions there are similarities: Reputation is based on people and includes their judgment and expresses an assessment [Peters, 2011].

As a relational construct, organizations cannot fully control their reputation. It can be seen as a social evaluation of a group of entities toward a single individual, a group of people, or an organization regarding certain criteria. More simply stated reputation is the result of what someone does, says, and what other people say about it [Gaines-Ross, 2008]. Thus organizations do not own their reputation; it is rather formed by the perceptions of others. Reputation cannot be *"enforced instrumentally, but only trustfully acquired"* [Zerfass, 2004]. Trust, the prerequisite of every corporation, is a critical factor for reputation. Although reputation is built upon trust, in turn, trust is an outcome of a sound reputation; these two concepts form a symbiotic relationship to each other [Picot et al., 2003; Ebert, 2009; Klewes & Wreschniok, 2009]. [Chun, 2005] considered an organization's reputation to be a synoptic standpoint of the perceptions held by all of the germane stakeholder

groups of an organization. A sound reputation sustainably strengthens an organization's position in the struggle for profitable clients in the hunt for talents, and in its affiliations with the stakeholders. A positive reputation takes the pressure of control and hence reduces costs, defines power in conflicts, legitimates imbalance in power attracts trading partners and office seekers [Eisenegger & Imhof, 2007].

In contemporary media societies, the media plays the central role in the process of forming reputation. Because of the propensity for scandals and negativity of the media, reputation is as fragile as it is important [Eisenegger & Imhof, 2007]. Besides the broadcast media, it can be strongly assumed that the Social Web becomes more and more important in the process of forming reputation. According to [Peters, 2011], the quality of an organization's reputation is influenced by the intensity of the stakeholders support potential. Thus from a business view, for an organization a good reputation is more than nice to have—it is essential. It significantly impacts the actions or behavior of stakeholders towards the organization. According to this, reputation is considered as an important—if not the most important—intangible asset of an organization. In this respect, it makes absolutely sense that executives ascribe an ever-greater significance to reputation management [Hexter & Bayer, 2009].

To methodically treat the diverse aspects of online reputation management (and analysis in particular), a design research approach is pursued. Design disciplines have a long record of building knowledge through forming artifacts (e.g. frameworks and prototypes) and the ensuing evaluation of its performance. Within this chapter the focus is on existing theory and research (i.e. literature review) of (online) reputation management. On that account, section 4.1 introduces reputation management as a holistic approach, whereby the entire process of reputation management is highlighted. Section 4.2 elaborates on online reputation management, the application of reputation management on social media elements. To analyze in particular online reputation, the process of online reputation analysis of an organization's reputation will be presented in section 4.3. Afterwards, section 4.4 reveals the incorporation of the fuzzy grassroots ontology in online reputation analysis. Finally, section 4.5 concludes this chapter with recommendable further literature.

4.1 The Process of Reputation Management

For an organization, a good reputation is, according to an economic perspective, an intangible asset used as sociopolitical legitimization to succeed in financial performance (e.g. liquidity and profitability) [Wilson, 2005]. The central role of this perspective is: Organizations with stronger positive reputations are able to attract more and better consumers (i.e. they are more loyal and buy broader ranges of products and services). Because the market believes that these organizations will deliver sustained earnings and future growth, they have higher price-to-earnings ratios, higher market values, and lower costs of capital. Moreover, in an economy where up to eighty percent of equity is derived from intangible assets that are difficult to assess, organizations are particularly vulnerable to anything that damages their reputation [Eccles et al., 2007]. Ninety-five percent of executives think of an organization's reputation as playing a crucial role in achieving their business objectives, sixty-three percent of organization's market value is ascribable to reputation, and the top ten world's most admired organizations (among them Apple Inc.) enjoy a total shareholder return of almost three times that of the five hundred largest US trading organizations [Beal & Strauss, 2008].

However, for stakeholders, reputation is information about an organization's reliability and goes as a distinguishing feature to other organizations. Some of the many stakeholders include affiliates (e.g. creditors, investors, and suppliers), communities (e.g. online and offline collectives), consumer (e.g. high-value, long-time and first-time customers), detractors (e.g. disgruntled employees or competitors), media (e.g. journalists and influential bloggers), opinion leaders (e.g. analysts, journalists, bloggers, thought leaders and executives of leading firms).

Reputation management is becoming a paradigm in its own right as a consistent way of looking at an organization and its performance. Commonly within business administration literature, it is assigned to the fields of marketing and Customer Relationship Management (CRM). Thereby the effects of reputation on the profit materialization are studied. However, within sociological literature (e.g. in the realm of media and communication), reputation management is most often associated with the fields of communication management and Public Relations (PR). At that, the question of how the organization's reputation is measured and influenced by (corporate) communication is studied [Peters, 2011]. Thereafter, usually insights about the significance of reputation for the integration and coordination of actions are converted into a business context.

According to the approach of integrated corporate communications an organization must present the same message to all of its stakeholders to convey coherence, credibility, and ethics. Communications operatives can help to build this message by combining the vision, mission, and values of the organization [Zerfass, 2004]. In the course of this, corporate communication

can be both internal and external [Röttger, 2005]. Internal communication can involve individual actors as employees, executives and board of directors, whereas external communication can encompass analysts, bloggers, collectives, competitors, creditors, customers, investors, journalists, leaders, suppliers, etc. Of course these actors can also be assigned to the previously presented stakeholder groups (e.g. affiliates, communities, consumers, detractors, media, and opinion leaders).

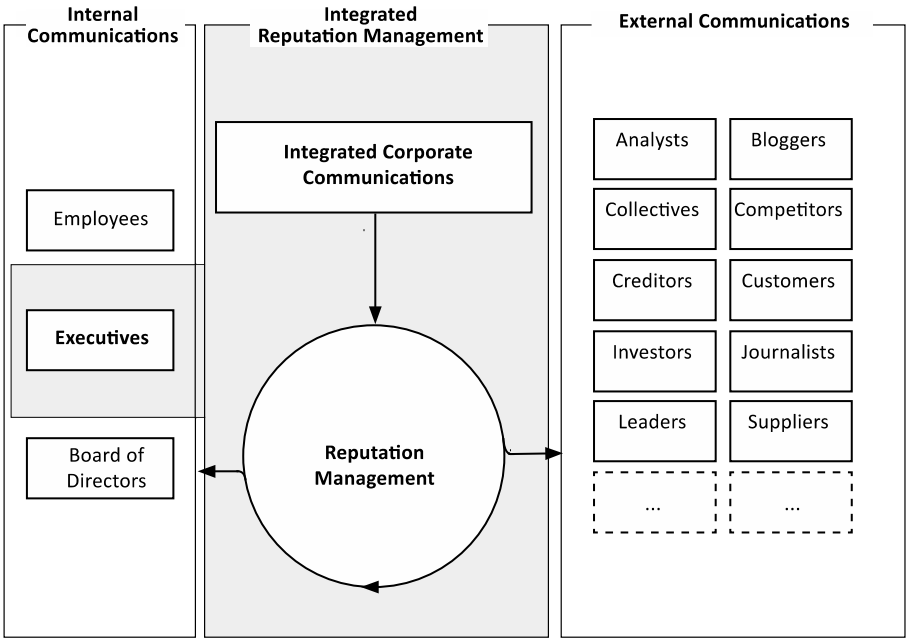

Figure 4.1: Integrated Reputation Management.

Though most executives know the value of reputation, it is also not uncommon for organizations to hire professionals to manage their reputation risks. According to [Eccles et al., 2007], effectively managing reputational risk involves assessing the organization's reputation among all germane stakeholders, evaluating the organization's real character, closing reputation-reality gaps, monitoring changing beliefs and expectations, and placing a particular executive in charge of these tasks. The assignment of this executive typically consists of tracking the actions of an entity and the opinions of other entities about those actions, reporting on the actions and opinions, and reacting to the report, creating a feedback loop. As a corporate officer position, this executive reports directly to the CEO. According to [van Riel & Fombrun, 2007], corporate communication is the set of activities required to manage and orchestrate all of the internal and external communications, which are aimed at creating favorable starting points with the stakeholders on whom the organization depends. It consists of the accumulation and dissemination of information with the common goal of enhancing the organization's ability to retain its license to operate [Zerfass, 2004]. Adapted from

[Burkhardt, 2009], in figure 4.1 the integrated reputation management and its actors are visualized.

Since there are not only one but numerous stakeholders, it may be possible that an organization not only have one but several reputations, depending on the stakeholders. Furthermore reputation can be decomposed analytically into three dimensions [Eisenegger & Imhof, 2007]:

- *functional reputation*: for competence,
- *social reputation*: for integrity, and
- *expressive reputation*: for attractiveness.

Thus, an organization's reputation is formed in the intersection of what an organization says about itself and the stakeholders' perceptions. The goodness of reputation is an indicator of the stakeholders' potential to support the organization in the future. Over time, the process of formatting reputation takes place among all stakeholders in a dynamic process.

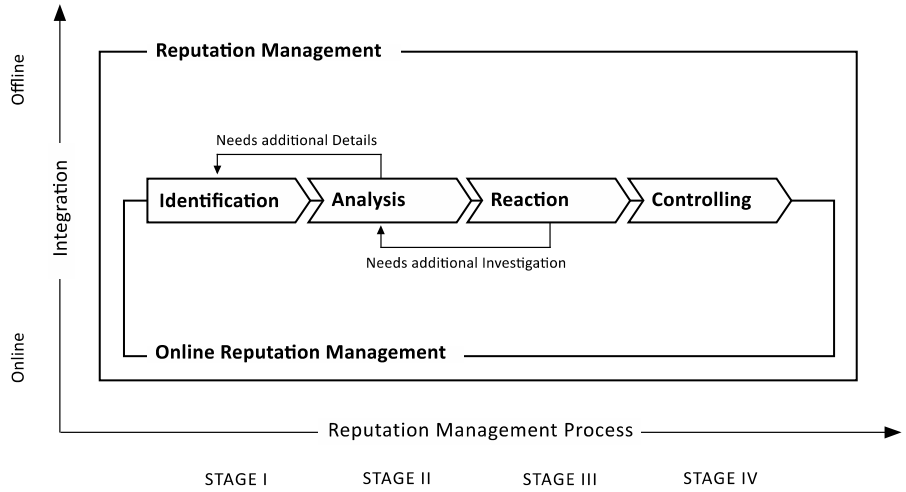

Figure 4.2: Process of Integrated Reputation Management.

As a whole, reputation management should come across as a holistic approach: It thereby includes the offline as well as online world. Due to the strong interconnectedness of these worlds and its increasing convergence, a separation is neither possible nor useful [Rolke & Köhn, 2008]. Figure 4.2 illustrates the process of integrated reputation management from a business administration viewpoint. It encompasses the steps of identification, analysis, reaction, and controlling. These steps are adapted from [Ingenhoff, 2004; Peters, 2011; Schreyögg & Koch, 2007], which use these steps for reputation as well as for issue management. Sometimes these authors include even more steps, but for simplicity reasons within this thesis the reputation management process consist of these four basic steps.

All the steps follow a business strategy for the setup, maintenance and expansion of a positive reputation. This management process involves and describes the actions of individual actors or the organization itself. In section 4.1.1 the challenging task of detection of reputation issues (e.g. threads and opportunities) is illustrated. Regardless of the amount of data and the cleverness of analytics tools, analysis is still needed. The sharpest analyst or most talented statistician is stymied without data, to be sure. Therefore section 4.1.2 is dedicated to the analysis (either by computer or human) of issues found. Section 4.1.3 reveals eventually necessary adaptations to the business strategy based on the analyzed issues. For optimizing communication with stakeholders, partial measurements in communication controlling should be considered. Section 4.1.4 clarifies this last but not less important step of controlling.

4.1.1 Identification of Reputation Issues

In consequence of the crucial role of the media (incl. the pre-media space of the Social Web; see chap. 2) it is most important to identify potential changes for a positive and risks for a negative reputation as early as possible. The pre-media space thereby encompasses the area between individuals, organizations, and (traditional) media. This area is not controlled by any of these single aforementioned entities but rather by all participatory entities together each on its respective ways. Therefore there is a link between the disciplines and processes of managing business relations' reputation and issue management. The new communication technologies play a key role in this interlinked process [Heath, 1998]. At this point clearly limitations turn up, since reputation issues often do not announce themselves through signals. Besides no universally valid rules can be defined how to search for issues. According to a particular business context, in fact, these criteria must be developed again and again. The perception must be extensively organized to be aware of the issues in an early stage. For multinational organizations (e.g. Apple Inc.) this holds all the more, because a fortiori national peculiarities must be considered in the different departments [Ingenhoff, 2004].

Reputation issue identification can interleave with the next step of reputation management, the analysis of the reputation issue. Consequently, after the issues are identified, a more precise analysis should follow. The next section illustrates the analysis in more detail.

4.1.2 Analysis of Identified Reputation Issues

In the analyzing step the found reputation issues need to be cleared and prepared for a business strategy. The exploration of various accompanying factors (e.g. fashion, scope and timing) serves as preparation. At this point an issue is classified and analyzed according to different dimensions to deduce an optimal valuation. Likewise, the stakeholders behind the issue have to be analyzed too. This analysis of the stakeholders includes an integration of their attitudes, intentions, plans, as well as their behavior. For example,

this can be acquired if the stakeholders are interviewed [Peters, 2011; Porák et al., 2007]. This direct and indirect interview includes the stakeholders' demands and expectations, as well as their perceptions, evaluations and assessments of their experiences with the organization, also in comparison to competitors. This ascertainment should basically be integrated using stakeholder interviews, Web monitoring, online reputation analysis, and reputation measurement in general [Fombrun & Wiedmann, 2001; Porák et al., 2007; Sterne, 2011]. Also to be considered are *"elsewhere gained insights"* about stakeholders such as in the context of CRM, or through activities such as online reputation analysis. In contrast to conventional stakeholder interviews, in the Social Web the stakeholders are self-initiators of their information. According to [Marti, 2011] this is significant because the organizations unintentionally can influence the stakeholder through their interviews. Yet, to increase effectiveness of individual activities, these activities should not only arise in an integrated manner, but should also complement each other. For example, the questions of direct stakeholder interviews can be supplemented or revised with the insights of online reputation analysis.

In doing so, the complexity increases by the problem of expectation of expectations. The derivation of business strategies adjusts itself not only on the analyzing step, but also take into account what is expected by the stakeholders regarding the actions of the organization self and other stakeholders [Beal & Strauss, 2008; Ingenhoff, 2004]. Put more simply, it is the stakeholders' opinions that matters, not what the organization thinks it should think about the organization. This results in a complex network of relationships of own and foreign expectations and actions. Therefore this step is to be regarded as a repeated iteration process. The permanent consolidation allows an adjustment and steadily improvement of the reputation management processes. It should be considered as a permanent feedback loop.

Since it is impossible to handle out of the analysis all issues coevally, a prioritization of the issues arises too. Different valuation criteria can be needed for such a prioritization. They are usually bound to the dimensions to parse, as well as to the classifications of the issues and stakeholders. In addition, portfolio techniques can help [Lucko & Trauner, 2004]. Thereby the issues and stakeholders are evaluated in a matrix. Even if the evaluation pursues an alleged objective strategy, these methods are based on a subjective evaluation and have to face the inherent problem that it could always be biased [Ingenhoff, 2004]. After a careful analysis of the reputation issues, reactions on it should be planned. The following section presents some possible reactions.

4.1.3 Reaction to Analyzed Reputation Issues

Depending on the analysis, the development and implementation of an appropriate business strategy is ensued. In the course of this the organization's reaction should be defined. As a result, the organization can communicate

with one voice [Ingenhoff, 2004]. Since several departments of an organization can be involved coevally in an issue, it is necessary that this strategy is developed with the inclusion of all concerned departments. Only then can a sound business strategy be developed and implemented. Following [Eccles et al., 2007], this strategy should be orchestrated by a particular executive with all the necessary power of decision.

By means of the insights about stakeholders, a starting point can be determined how and to what extent the organization should change its actions and services to improve its reputation to the different stakeholders. In other words, the insights are non-tolerated to remain exclusively in the department the information were collected and integrated. Rather, they must be forwarded to all concerned departments. Subsequently it must be decided whether and how to alter acting. On this basis, communicative actions can be planned. This can, in turn, generate information about the actions and performance of the organization.

A combination of variables and characteristics can serve as basis for a specification of strategies and handling of the issues [Beal & Strauss, 2008; Ingenhoff, 2004]. For example, it can be distinguished between an active, adaptive, proactive and interactive handling of an issue [Ingenhoff, 2004; van Gaalen, 2009]. They are often not selective distinctions; some of them may overlap. If the organization, for example, decides to take on a proactive strategy and in this manner attempts to actively react on the development of an issue, the organization must enter into a dialogue with the corresponding stakeholders and communicate with them. In the end, what strategy an organization picks depends on both (only partially influenceable) external as well as internal factors.

Same as the step that leads to a business strategy, the implementation of this strategy is not yet systematically studied in existing literature. In order to influence the trend of an issue, mostly possible arrangements are presented. As a result the chosen strategy can tend inward or outward. The implementation of the business strategy comprises communicative arrangements such as active participation in discussions, the dissemination of press releases, and installation of a board of experts, campaigns or lobbying. Likewise important are internal arrangements as for example changes in production policy. Here too, the integration of various departments strikes the eye. Again, a particular executive should orchestrate this as well. This executive should control the whole reputation management process. This is presented in the following section.

4.1.4 Control the Reputation Management Process

For a continuous improvement and legitimization of the reputation management after it, an evaluation is of utmost importance. This evaluation should not only take place at the end as a result-control, but rather throughout the whole process of an organization's reputation evaluation itself. Some

authors even state a separate assessment of all individual steps within the process of reputation management to reveal inherent problems and to detect difficulties in the transition to the next step on time [Ingenhoff, 2004]. However, coherent constructs to integrate reliable measure and evaluate the results, are to a great extend deficient yet. This is certainly also related to the fact that a concrete definition of reputation is missing so far.

Various kinds of calculations of an occurred reputation loss, as a result of poor management, are in the cards. Particularly for a successful implementation of reputation management, which is reflected in the fact that no or only limited reputation loss is to be expected, a corresponding quantification has been modeled by [Fombrun & Wiedmann, 2001]. Their reputation quotient compares an organization's public image with its self-image. The latter is based on interviews with employees and executives, whereby the former is determined through interviews of relevant stakeholders, or through media analysis [Porák et al., 2007]. [Eisenegger, 2008] presents his reputation index that can adopt values between -100 (i.e. exclusively negative) and $+100$ (i.e. only positive). These measurements represent relatively simple and convincing methods for measuring reputation.

4.2 Online Reputation Management

Word Of Mouth (WOM) communication is an important purchasing decision input. People trust most those who share similar interests, have the same occupations, and hold akin political convictions [Beal & Strauss, 2008]. They heavily take into account the options of like-minded people. As a result like-minded people have a vital influence on consumers purchasing behavior [Heider, 1946; Lin, 2002]. Likewise more and more consumers get to know the structure of the Social Web, so they share evermore their experiences with organizations online. Social media elements are shifting the way in which these consumers communicate by giving them the opportunity to contribute to discussions about anything. Through these contributions the consumers (in the collective) turn into producers of an organization's reputation. In a convergence culture the role changes from consumer to prosumer [McLuhan & Nevitt, 1972; Toffler, 1980; Jenkins, 2008]. Hence these prosumers are amplifying voices in marketplaces and exerting far-reaching effects on the ways in which other consumers buy.

Up to now, organizations have often asked customers to trust them. In a prosumer-oriented market, the shoe is on the other foot and organizations themselves must show that they trust the consumers. Yet, whether an organization trusts them or not, they will continue to talk about the organization in reputation-increasing or reputation-detracting mode. Listening with respect will give an organization interesting insights about its online reputation. This online reputation has implications for organizations and should be taken seriously while doing (electronic) business. Here electronic is written in parentheses to indicate that with social media it is no longer sage to dif-

ferentiate between online and offline business. The same holds for reputation. Within this PhD thesis the focus is on online reputation management. Although reputation management is a holistic approach, through the Social Web it is becoming more and more important.

According to [Meier & Stormer, 2009], electronic business means the exchange of services with the help of media to achieve added value. In electronic business, all stakeholders of an organization can be prosumers too, and the relationship therein generates added value for all involved. This relationship may take the form of either a monetary or an intangible contribution. A central need of electronic business is to appropriately manage the organization's relationships with its consumers [Bruhn, 2002]. This includes reputation as well. As the Social Web is not moderated or censored, users can say anything they want, whether it is good or bad. This freedom indicates the need to manage these relationships by carefully watching and, if necessary, interacting with them in an appropriate way [Scott, 2011]. Because there are plenty of examples of how not to interact, this communication should be carefully considered and eventually relinquished to employees specially trained for the Social Web to optimize business relationships [Portmann, 2008]. Through participation in the conversations, the affected parties can improve the organization's image. Increasingly, organizations are looking to gain access to conversations and to take part in the dialogue. Mainly in electronic but also in traditional business, cautious monitoring of the organization's online reputation should be considered.

Online reputation management is the task of monitoring, addressing, or rectifying undesirable or negative SERPs or mentions in online media. A SERP constitutes the listing of webpages returned by a Web search engine [Hearst, 2011]. An example is Google's hit list (e.g. answering the Apple search from chapter 2). The objective is to have stakeholders see positive mentions of an organization's brand and speak about it in a positive sense by achieving and maintaining a positive online sentiment. Organizations often use Search Engine Optimization (SEO) to increase an organization's website ranking on specific keyword's SERP. An online reputation management strategy monitors online buzz and sentiment about the organization by engaging stakeholders positively. This portends that not only offline but also online dimension should be included and covered. Following [Eccles et al., 2007], for effectively managing this particularly online dimension, an executive should be responsible for directing all Web activities.

Especially when the online dimension of an organization is of importance, this executive should be located close to the Web activities. Hence, this executive should have on the one hand the CEO's ears and on the other hand a deep and broad understanding of the Web. [Meier & Zumstein, 2012] assign such tasks to the Chief Web Officer (CWO), in the contrast to traditional reputation management, where these tasks are more probably assigned to the Chief Communication Officer (CCO). The CWO is the highest-ranking corpo-

rate officer (executive) in charge of an organization's Web presence. The assignment of this executive typically consists of tracking the actions of an entity and the opinions of other entities about those actions, reporting on the actions and opinions, and reacting to the report, creating a feedback loop. Coevally the relationships with all stakeholders will be encouraged and expanded by using social media elements (and applications). Furthermore, for optimizing communication, particular measurements in the communication controlling should be upgraded to an online dimension. This upgrade should apart from the organization's website necessarily also include the pre-media space of the Social Web.

On that account additional financial, technical, and computational resources (e.g. hardware and software) could be necessary. Likewise important is that explicitly for the online division additional employees are at hand. This could be communication operatives, corporate bloggers, social media managers, Web administrators, and so forth. Within this PhD thesis they are summarized as communication operatives. Indeed this is only an essential but not a sufficient resource and precondition for a successful online reputation management. Important is that the skills and knowledge of handling is tailored to the Social Web. Thereby it is essential that the monologic one-way understanding of traditional online communication is widened to a dialogue-based one to use the full potential of social media elements for reputation management. Yet, without the willingness and permission of a participatory and interactive communication with the stakeholders this is impossible. Both, the dialogical understanding of communication and the willingness and permission to partake imply that an appropriate corporate culture is borne by its management.

From the perspective of stakeholders the impacts and changes to an organization's reputation is severe [Peters, 2011]. Published content (e.g. photos, videos, contributions to discussions, ratings of products, etc.) is visible and accessible to everybody by most of social media applications. Communication can be stored permanently and as a result it can be retrieved at any time. Web archives arrange so that this also holds even if the content were allegedly deleted. Likewise in the majority of cases, data published to the Web are findable by Web search engines and reproduction is enabled by the digital nature of the available data.

Nevertheless, characteristics like the power to upload information, service-oriented design, open API and technologies such as the APP and RSS fosters to a great extend an acceleration of dissemination of information (e.g. comments, references, ratings, etc.). All of the following social media elements can contain these kinds of information, referred to as an integrated platform. In the following sections the implications of stakeholders' conversations on the organization's reputation is illustrated: Section 4.2.1 comments on blogs in the process of online reputation management; section 4.2.2 microblogs; section 4.2.3 folksonomies; section 4.2.4 wikis; and last but not least section

4.2.5 on social networks. At the very end, section 4.2.6 points up some possible reactions to all of these social media elements in online reputation management process.

4.2.1 WEBLOGS

Stakeholders can actively avail their own blogs to publish personal or mediated opinions, experiences, ratings, and estimations. They can also take up Web content, reporting, discussions (e.g. in the blogosphere or in forums) and offer a particular opinion on that in their blog. Likewise the stakeholders can comment blog posts or they can start off or carry on discussions on a blog operated by other stakeholders. Blogs can go for the organization of actions such as protests or boycotts, for example. All in all, using blogs, stakeholders open up a possibility to communicate to other stakeholders and in doing so they become able to act in a much wider range. Blogs allow the communication among stakeholders. Because of the spatial and temporal shift, this was not possible as quickly in the past. Blogs expand communication that also other unknown stakeholders from diverse social networks can communicate together [Portmann & Hutter, 2011].

Blogs cannot only be used actively but also passively. In this sense, they act as an information source. Hereby stakeholders provide, independent of the organization and the media, information about organizations. On the other hand organizations provide direct information through their corporate blog. Finally, also the media provide information about organizations in their blogs [Peters, 2011]. The ruthless candor in respect to different issues bestows blogs on additional information source. These issues can be special subjects as well as unknown branches that are (yet) not heard in traditional (mass) media (e.g. print, radio, television).

According to [Peters, 2011], media often use blogs actively to expand their offering and reporting. Thus, the feedback possibilities are of utmost relevance for an organization's reputation. These feedback possibilities allow media to recognize their resonance and generate further information related to a juicy, high-interest subject. Furthermore blogs are relatively independent of fixed submission deadlines. That allows a faster dissemination of information. Through the personal and authentic style inherent to blogs, they offer good publishing opportunities that stimulate further discussions. In turn, these discussions may affect an organization's reputation again. Comparable to stakeholders, journalists passively avail blogs as an alternative source of information [Hächler, 2010]. This allows fervently discussed posts to spill over to traditional media. In this way, hot discussions can go into the offline world and thereby reach even more consumers [Portmann, 2008]. This can be relevant for an organization's reputation such that the take up by the traditional media reaches a massive increased coverage.

4.2.2 MICROBLOGS

Microblogs can either be used to publish personal and intermediary perceptions, experiences, and ratings or as reference to Web content. Normally microblogs are used to inform members of a (personal) social network (e.g. friends or followers). Thereby they can be used as a passive source of information for both personal and independent information on organizations as well as direct information from the organization [Portmann & Hutter, 2011].

Often the media avail microblogs actively for the dissemination of short news in the form of news tickers [Peters, 2011]. Thereby mostly the media-owned information source (e.g. news portal, weblog, etc.) is emphasized and linked. Through microblogs media makes its readers aware of as relevant considered reporting. Within social networks the dissemination of information can be strongly accelerated and expanded [Portmann & Hutter, 2011]. Moreover often by journalists microblogs are used passively as a source of information, for example on current events [Hächler, 2010].

Due to its constrictions (e.g. to only 140 characters), posts can be facilitated only with a limited substantially body. Their effect on an organization's reputation is rather small [Peters, 2011]. However, microblogs may indirectly initiate, accelerate, and expand dissemination of information. In the course of this a more detailed discussion in other locations such as in traditional blogs or in the traditional media can go adrift [Portmann & Hutter, 2011].

4.2.3 FOLKSONOMIES

Social bookmarking platforms primarily go for saving websites as bookmarks that contain information about an organization's acting and services. By an active usage for a given situation the information has priority. Due to the collaborative nature of social bookmarking platforms by collaborative tagging of Web content, stakeholders can use these platforms for information finding. With respect to an organization's reputation this is important because social bookmarking rests on experiences of other users and thereby presents possibly more specific information than traditional Web search engines [Peters, 2011].

Depending on the media content sharing platforms constitute an alternative information source for perceptions and experiences with an organization. Social bookmarking platforms are limited to linking Web content of such platforms (i.e. active usage) together with the possibility to use them as alternative search engine (i.e. passive usage). Both could be a thread to an organization's reputation. Beyond that, media offers its users the possibility to merge reporting with Web portals. In this manner they promote an accelerated and extended dissemination of information within social networks [Portmann & Hutter, 2011; Peters, 2011].

4.2.4 WIKIS

Stakeholders can actively exert wikis, and especially Wikipedia to publish reputation-related information, or edit existing posts under this objective. This may be relevant for the reputation building process, because the chances that apart from more or less universal information, in Wikipedia also critical information can be found [Peters, 2011]. Accordingly, this information does not increasingly pass into oblivion. To change something in Wikipedia is relatively awkward [Fuchs, 2010]. Compared with traditional encyclopedias, the information in Wikipedia is frequently updated. From this point of view the information has priority. Wikis (and Wikipedia in particular) are used as a source of information. Information published on Wikipedia get a relatively large audience. Depending on individual user assessment of the quality of Wikipedia contributions, there could be an impact on an organization's reputation. With Wikipedia's rate this page function, the trustworthiness of Wikipedia entries should boost and thereby critical information can constitute a higher risk to organizations.

A small and closed group of administrators only controls Wikipedia, as an empirical network analysis has shown. Even though every user has the opportunity to write articles, most of the articles are written by a small number of authors [Stegbauer, 2009]. Nevertheless, organizations should be aware of articles about critical issues because of Wikipedia's wide reach. According to [Hächler, 2010], Wikipedia is often used during searching by online journalists. In these premises and with regard to media's active use, a clear statement cannot be made. Wikipedia—as the most popular wiki—may be relevant to the reputation building process, if journalists, for example, use therein published information unverified in their coverage, and in turn affect a disperse audience with potentially false information [Peters, 2011]. Again, by the rate this page function, false information should decrease.

4.2.5 SOCIAL NETWORKS

If a social network profile exists and contacts therein are available, then the stakeholder can use this network for an active dissemination of information (e.g. content, experiences, opinions, and perceptions). According to [Portmann & Hutter, 2011] this may be relevant for an organization's reputation if the network is rather large and the information finds its way into other social networks. This already points to the passive possibility of usage of social networks. Solely through the registration with a social network, a user often gets information about other members (i.e. friends) in his social network by push-function, without this was being actively sought after. Overall it seems that social networks increasingly are becoming central points of organizations. This means that the stakeholders gather their online activities in these profiles and make it widely available to their contacts [Peters, 2011].

Normally media allows its audience the dissemination of reporting within a network by social media buttons (e.g. like or +1 button). By pressing these buttons the coverage are disseminated within a stakeholder's network. For journalists networking platforms offer to establish personal contacts to informants, for example. Thereby they can acquire information through the membership in different networks [Portmann & Hutter, 2011].

4.2.6 Interaction with Social Media Elements

For a modern and effective online reputation management rules must be modified and extended. This obliges a synchronization of individual responsibilities of online reputation management in the corresponding departments as well within corporate communication itself. On the other hand, existing rules, for example in dealing with customers or journalists, should be completed by Social Web guidelines. Based on these modified and expanded structures (i.e. rules and processes) a strategy for online reputation management can be developed. This strategy begins with the planning of activities to build, maintain and expand an organization's reputation.

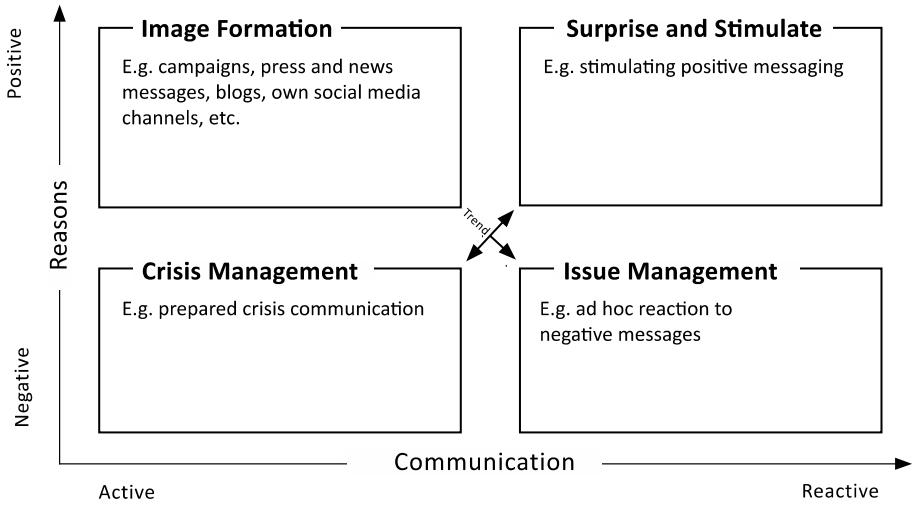

Figure 4.3: Possible Interaction with Social Media Elements.

An organization is increasingly at the mercy of its stakeholders. Adapted from [van Gaalen, 2009], in figure 4.3, a summary of different possibilities for online reputation management is presented to help optimize an organization's online reputation. As at the beginning of this chapter (see sect. 4.1) online reputation management was introduced, located in the interface between marketing, CRM, corporate communication, and PR. At a strategic level the CWO is responsible for a sound online reputation. As demonstrated, it is increasingly becoming easy for stakeholders to express their opinions about an organization online and share them with others. UGC presents organizations with opportunities, as well as threats, for actively creating a positive reputation and reactively responding to stakeholders' messages.

On one hand, communication has either a positive or negative origin, and, on the other hand, there is active or reactive communication. Active means that an organization pushes positive influence ahead: reactive, however, only responds to culled Web content [Kaiser, 2011]. Resulting from this are four categories of online reputation management [van Gaalen, 2009]:

- *Image formation*: The goal is to build a presence and to take an active part in those online places where stakeholders are to be found. All this online communication is geared towards proactively informing and engaging stakeholders in order to create a positive reputation [Peters, 2011]. The organization may, for example, mount campaigns in the pre-media space by publishing press releases, writing blogs, news items, and so forth. These are the most common ways of building an online reputation and are normally sender-oriented. These campaigns, which are generally found on organization's corporate websites, can also be supplemented by placing content on external websites in pre-media space.

- *Issue management*: The main focus here is to reactively handle negative mentions (i.e. track down complaints and negative messages). It is important to respond adequately to social media elements [Portmann, 2008]. Thereby it is to ensure that the negative messages end up lower in the search results that is. When facing sensitive issues or unexpected negative reporting, SEO (i.e. form of improving the visibility of benevolently minded website) is an option.

- *Crisis management*: Many organizations have crisis plans, but unfortunately, not many have woven social media elements into their plans [Beal & Strauss, 2008]. In general, if an organization communication operative communicates quickly, truthfully, and transparently, then the organization is in a good shape. The three-word mantra is sincerity, transparency, and consistency. It is increasingly important for an organization to be transparent and communicate clearly what their role or point of view is during certain crises [Thiessen, 2011].

- *Surprise and stimulate*: Finally, there are the positive reactions to reporting about an organization. Organizations are becoming increasingly aware that they should not only have a focus on negative reporting but also on positive reporting [van Gaalen, 2009]. The positive serves not only to counteract negative reporting but also to stimulate and reward these ambassadors of the organization. These brand ambassadors are extremely precious, since they influence other stakeholders with their positive reporting about the organization.

Online reputation management is more than just focusing on negative reporting and setting things right again. It is about monitoring and managing both negative and positive reporting in an active or reactive way. In other words, this means that even without an active online engagement, a success-

ful and modern reputation management requires additional and modified structures in order to keep an organization able to act. Many organizations make the transition from actively positive communicating to reactively negative communicating. But to create a positive online reputation, there is no way organizations can avoid using the other elements of online reputation management as well. The business landscape has become a business conversationscape. It is a connected world after all. By ensuing periodic controlling of reactions (see sect. 4.1.4), the online reputation management process closes. So for an organization it may be essential to put its measurements accordingly. However, so far, there are only few measurements for social media controlling [Sterne, 2011].

4.3 Online Reputation Analysis

The Social Web consists of software that provides online prosumers with a free and easy means of interacting or collaborating with each other. Consequently, it is not surprising that the number of people who consume Web content at least once a month, for example, has grown rapidly in the past few years and is likely to increase further in the foreseeable future. These Web contents strongly influence what people think about organizations [Donges, 2008] and what products they purchase. The influence on potential purchases is leading many organizations to strategically conduct online reputation analysis. Through this analysis, it is possible to identify conversations that mention entities of the organization. Through participation in the conversations, the affected parties can improve the organization's image, mitigate damage to their reputation posed by unsatisfied consumers and critics, and promote their products.

In the literature, several approaches are described how to identify online reputation; most of them rely on the management task of online reputation analysis [Fombrun & Wiedmann, 2001; Eisenegger & Imhof, 2007; Ingenhoff & Sommer, 2008]. Nevertheless, the significance of these analyses is critical considering that a negative SERP will often be picked first when listed with an organization's website. Imagine the blogger Leontien Aarnoudse bashes on Apple Inc. for their inhuman working conditions at a Chinese supplier of iPad parts [Aarnoudse, 2011]. Now then, when his post appears on (the first page of) Google's hit list together with Apple's official website, many users of Google clicks the negative hit before anything else. Hence following a pursuant strategy, organizations evermore try pointedly to impinge on their reputation by planning, organizing and implementing, as well as controlling and analyzing of actions to be taken. By trying to influence the opinions of its stakeholders, corporate communication makes a contribution to this. The acting of reputation management will thereby be both enabled and constrained by structures. They, in turn, are reproduced and modified through actions; thus emerges a cycle. Crucial for a successful reputation management is that these structures are tailored to the new and changing possibilities of information in the Social Web. If such adaptations remain undone,

then the scope of the reputation management is limited and the chance of a successful exertion of influence shrinks.

To proactively shield their reputation from damaging content, organizations increasingly rely on online reputation analysis. Because UGC has enhanced the public's voice and made it very simple to make articulated standpoints and, given the advances and attractiveness of search engines, these analyses have recently become more important. Recall [Beal & Strauss, 2008]'s findings that people trust most other like-minded people for credible information in addition to that (see sect. 4.2). They can map opinions and influences on the Social Web, simultaneously determining the mechanisms of idea formation, idea spreading, and trendsetting. With its emphasis on influencing SERPs to protect a organization, online reputation analysis can be viewed as a field that relates to other areas of online marketing, such as SEO [Dover, 2011] and WOM marketing [Silverman, 2011]. WOM marketing is an unpaid form of promotion in which satisfied customers tell other people how much they like a business, product, service or event (see also surprise and stimulate quadrant in fig. 4.3).

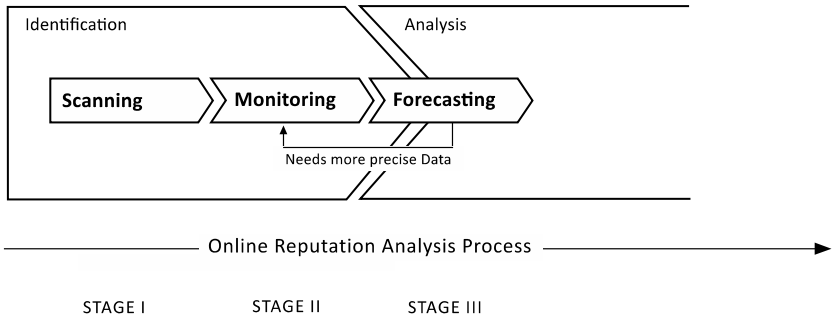

Figure 4.4: Possible Interaction with Social Media Elements.

Figure 4.4 illustrates the single steps of online reputation analysis. A challenge thereby is the prevention of flooding caused by vast amounts of data. Scanned issues must be summarized into manageable topics, and their changes must be surveyed in the ensuing permanent monitoring to avoid surprises. Monitoring is a method of reputation analysis that is equivalent to scanning but watches a selected range of topics only. In a sound online reputation analysis also a forecasting should have its place. In the following section 4.3.1 first the scanning for online reputation issues is explained. Then in section 4.3.2 the monitoring step of online reputation analysis is elaborated followed by the forecasting of online reputation issues in section 4.3.3.

4.3.1 Scanning of Online Reputation Issues

As reported by [Ingenhoff, 2004], the goal of scanning for an organization's reputation, on one hand, is the early detection of changes in the environment of the organization that may affect or restrict the organization's scope. On

the other hand, new sectors can be brought to light through scanning. To position itself as an expert and opinion leader and to realize new opportunities, the organization can occupy these new sectors. Another goal of this approach is to evaluate the reputations of competitors; occasionally, a competitor will launch an unknown product or a new production method that can be detected through scanning [Portmann, 2009; Portmann & Meier, 2010]. Now, before an organization should start any online reputation analysis campaign, it is important to identify which brands and identities could come under fire [Beal & Strauss, 2008; Sterne, 2011]. An organization's name, brands, service, products and executives' personal reputation are a given — communication operatives certainly want to know any time these are mentioned. Likewise an effective analysis also requires keeping a watchful eye on competitors and trends in the industry. To ensure you are not blindsided by a reputation crisis it is furthermore recommended to carefully observe marketing campaigns and known weaknesses. Certainly, these analyses, again, incorporate offline as well as online world.

Scanning refers to the (more or less) undirected and therefore inductive monitoring and examination of an environment. However, often there is a lack of systematic analysis and concatenation with the target system of an organization to scan and select issues. To avert crises, the focus is then only on the potential risks coming from issues [Ingenhoff, 2004]. Yet, taking into automatic text content analysis systems, an improved processing of Social Web data is anticipated. According to [Sterne, 2011], this automatic text content analysis may include a sentiment analysis or more sophisticated forms of language's syntax interpretation. The intention is to determine someone's attitude or the overall tonality with respect to some topic and depends often on Natural Language Processing (NLP). However, the interpretation of these data into meaningful information cannot yet be fully replaced by computers, since the interpretation depends strongly on human sensitivity, experiences and the power of finding associations. Nevertheless, the applications of computers can lend valuable assistance in the scanning process for Web data. A smartly detection of dangerous information is a competitive advantage for an organization and can act contrary to a loss of reputation.

The aim of the scanning results in, on one hand, an early detection of developments in the internal and external environment, which could possibly affect the scope of the organization. On the other hand, new area for the organization can be detected and occupied in order to take a position and realize opportunities. The challenge of scanning primarily lies in the prevention of information overload with information, whose impact cannot be assessed with certainty. This information has to be selected and condensed into manageable issues. To avoid unexpected emergencies, their changes have to be observed permanently in subsequent monitoring. The process of selection is essential for the further management of the issues.

4.3.2 Monitoring of Scanned Online Reputation Issues

In contrast to scanning, monitoring is a deductive task. Mainly with regard to possible changes, monitoring refers to the continuous and targeted observations of already identified reputation issues that are marked as important. It is not only assigned to Social Web data collection, but supervises the entire process. The instruments are considerably identical to those of scanning, with the difference that in this step they are targeted to observe a preselected range of issues. For an observation of known issues, external service providers, media database searches and report feeding-in systems can offer plausible solutions too [Ingenhoff, 2004; Peters, 2011].

The focal point is the analysis of social media elements, which can have a great impact on the reputation of an organization since it is impossible to control Social Web topics and their positive or negative public awareness. However, nowadays information can flow from Social Web to the pre-media space and then can be admitted by the traditional media. Hence, monitoring must support the pre-media space's increased influence [Peters, 2011]. The most obvious source is often neglected: the organization's employees. To integrate them into the monitoring process, in practice can prevalently result in added value for an organization. The next step is a forecasting of a reputation issues.

4.3.3 Forecasting of Identified Online Reputation Issues

To forecast trends and events that can yield business-related issues, scenario techniques, Delphi methods, cross-impact and trend analysis can be used [Ingenhoff, 2004]. Chronologically, the step of forecasting can also be used before the monitoring process in order to analyze potential issues and observe them specially. Based on history, forecasting techniques assume that trends can be predicted. Since changes may often take unexpected progressions, these predictions are imprecise and solely point roughly into a direction of development of an issue. Many organizations already use tools to locate and forecast reputation issues, but not only the early identification but also their timely processing by the organizations is important. Since in online reputation management a quick reaction is expected, in the task of forecasting also the step of analysis is in some cases associated. Appropriately trained communication operatives forecast and analyze the scanned and monitored issues: thereby forecasting and analyzing can comingle.

With the rising use of the Semantic Web technologies, thus forecasting could be supported, for example by including Web analytics (e.g. Google Analytics[27]) into the forecasting process [Meier & Zumstein, 2012]. Assembled information could be automatically or semi-automatically analyzed. This information could among others stem from log file analysis (e.g. parsing a log file from a Web server), page tag analysis (e.g. recording what users click while browsing) or sentiment analysis (e.g. extracting subjective information of users). Based on these analyses, for example, rules can be inferred

(e.g. by humans or computers). With the Semantic Web-inherent RIF layer, these rules can be transported and integrated into a forecasting system of online reputation issues.

This additional information can then also support implemented scenario techniques, Delphi methods, cross impact and trend analysis. Consequentially, it can help communication operatives to make better decisions and to form more accurate forecast trends. This is a possible field of application for discrimination, for example (see chap. 3). Discrimination is the machine learning task of inferring a function (e.g. rule) from supervised training data (e.g. the mentioned analysis) by the help of communication operatives which take corrective actions. Within this PhD thesis, however, this is not a topic warranting further discussion.

4.4 Use of Ontologies for Online Reputation Analysis

As already introduced in chapter 2, ontologies provide a shared vocabulary of a domain (i.e. the stipulation of objects and concepts, and their properties and relations). To support communicative operatives in the online reputation analysis endeavor, such an ontology can be consulted. On this basis, in the online reputation analysis process, the communication operatives can spot relationships or concepts, which were not known before. Thereby also new sectors could be brought to light.

For communication operatives to be supportive in the scanning process, the ontologies should be represented in human-understandable form. Therefore section 4.4.1 introduces the opportunities of interactive visualization (e.g. using Topic Maps) for the communication operatives to help to detect related hotspots. These Topic Maps rely on the fuzzy grassroots ontology introduced in chapter 3. A fuzzy grassroots ontology-powered interactive Topic Map allows improving the ability to detect early changes in an organization's environment. This is presented in section 4.4.2. Based on a fuzzy grassroots ontology, for example, new products of competitors can be sensed in a timely manner. In the same way the fuzzy grassroots ontology supports the scanning process, it can also facilitate the monitoring process. The repeatedly updated fuzzy grassroots ontology over time reveals, for example, changes in user classification of products (e.g. own and competitors) or steadily minimizes data loss and repetition. It is furthermore possible that in future the fuzzy grassroots ontology backs communication operatives in forecasting using the Semantic Web technology (e.g. RIF).

4.4.1 Exploration through Interactive Visualization

In order to achieve all the stated challenges in a simple manner, visualization techniques should empower communication operatives to spot patterns in Web content, identify areas that need additional analysis, and make sophisticated decisions based on these patterns [Zudilova-Seinstra et al., 2008; Hearst, 2011]. The human capability to converse, communicate, reason, and

make rational decisions in an environment of imprecision, uncertainty, incomplete information, and partial truth is supported by this visualization. Furthermore through visualization the information overload can be leveraged out. The manner in which communication operatives experience and interact with visualizations affects their understanding of the data; they benefit from the ability to visually manipulate and explore. This can lead to information sources from which knowledge can be derived. Likewise, visual interaction can support gut instincts and, confronted by management, provide an instrument to both substantiate theses and support viewpoints.

Besides mere visualization, an interesting feature of this method is the ability to discover hotspots through interactive navigation possibilities [Hearst, 2011]. To increase the ability to explore the data (and thus, to better understand the underlying context of social media elements), an effective integration of the visualization and interaction applications is important. According to [Ward et al., 2010], interactive visualization can be used at each step of knowledge discovery (i.e. the process of automated reputation issue mining for characterizing underlying patterns). Nevertheless, the field of analyzing data to identify relevant concepts, relations, and assumptions, combined with the conversion of data into computer language, is known as knowledge representation [van Harmelen et al., 2007; Weller, 2010]. The fundamental goal of knowledge representation (and reasoning) is to represent knowledge in a manner that facilitates drawing conclusions [Ward et al., 2010]. In other words it analyzes how to use symbol systems to represent a domain of discourse, along with functions that allows formalized reasoning about objects. Yet, because knowledge is used to achieve intelligent behavior, the fundamental goal of knowledge representation is to present data in a simple way. In the field of artificial intelligence, problem solving can be facilitated through a reasonable selection of knowledge representation [Orio, 2010; Sirmakessis, 2005]. Presenting context in the right way makes certain problems easier to solve.

According to [Spitzer, 2000], to at best understood, knowledge should somehow be organized in the same way that it is represented in the human mind or at least in the form of human language. This is indicative of NLP. Natural languages are capable of enunciating everything that can be stated in any artificial language with the same level of minutia and rigor, but can also put up with a degree of imprecision. In contrast, artificial languages are valuable because they do not tolerate imprecision, but what they claim to be so clear-cut may have no relation to what is intended. Various notations for logic are designed to represent the final precise stage, but most fail to provide intermediate forms that can bridge the gap between a vague idea and its formalization. Here fuzziness fares well, as it is derived from the human capability to perform a wide variety of tasks without any measurements or computations [Zadeh, 2004].

Recent developments in knowledge representation have been driven by the Semantic Web, and include development of XML-based knowledge representation languages and standards, such as RDF, RDFS, and OWL [van Harmelen et al., 2007; Weller, 2010]. However these languages currently rely largely on formal logic, which average media users, such as communication operatives, typically cannot adopt. Developments in the Social Web for a less formal visual knowledge representation are, for example, Tag Clouds. These are visual representations for data, typically used to depict tags on websites. Figure 4.5 illustrates an example of such a Tag Cloud. This format is useful for quickly perceiving the most prominent terms and for locating a term (e.g. alphabetically) to determine its relative prominence [Halvey & Keane, 2007]. When used as knowledge navigation in the Web, the tags are linked to associated entries. Yet, only humans can derive knowledge from them, since for computers Tag Clouds yield insufficient information. However, there are various efforts going on to cause Tag Clouds to be more computer-usable [Hassan-Montero & Herrero-Solana, 2006; Kaser & Lemire, 2007].

Figure 4.5: Tag Cloud Example.

Further approaches of knowledge representation are for instance in business settings Topic Maps [Pepper, 2010], in generic software programming Concept Maps (e.g. the Unified Modeling Language UML) [Novak & Cañas, 2006], and in psychology Mental Maps (e.g. Mind Maps, Cognitive Models, or Mental Models) [Kitchin, 1994].

Topic Maps (see fig. 4.6) are related to Concept Maps (in that both connect concepts or topics via graphs), while both can be compared with Mental Maps, which are in many cases limited to hierarchies and tree structures. Among the numerous representation techniques for visualizing ideas, organizations, processes, Concept Mapping, is exclusive in philosophical ground, which makes concepts, and propositions made up of concepts, the main components in the structure of knowledge and construction of meaning [Novak, 2010]. Another contrast between Concept and Mental Mapping is the speed and freedom when such a map is created. A Mind Map, for example, reflects what someone thinks about a single topic or can involve group brainstorming. A Concept Map can be a map, a system view, of a real (abstract) system or set of concepts. Concept Maps are more free practice, as multiple hubs and clusters can be formed, unlike Mind Maps, which fix on a particular conceptual center.

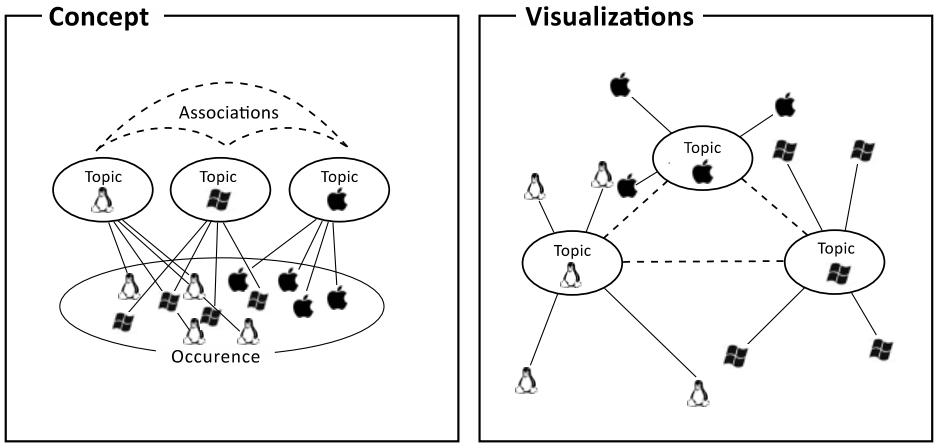

Figure 4.6: Topic Map Concept and its Visualization.

Topic Maps are to some degree related to RDF, where, however, the former are centered on topics while the latter on resources [Portmann et al., 2012]. RDF is based upon the idea of making statements about Web resources in the form of triples (see chap. 2). Topic Maps on the other hand are not limited to triples, and they represent information using a topic (representing any concept), association (representing hyper-graph relationships between topics), and occurrences (representing information resources). Furthermore, while RDF directly annotates resources, Topic Maps create a semantic network layer (i.e. a virtual map) above the information resources, leaving the information resources unchanged. Topic Maps explicitly support the concept of identity merging between multiple topics or Topic Maps. Furthermore, because ontologies are Topic Maps themselves, they can also be merged, allowing the automated integration of information from diverse sources into a coherent new Topic Map.

On one hand, the fuzzy grassroots ontology provides computers with a general knowledge of vague human concepts. On the other hand, the fuzzy grassroots ontology-based and interactive visualization of this knowledge helps communication operatives to find related patterns. The communication operatives can be backed by the fuzzy grassroots ontology by querying the Social Web for example. Thereby the interactive visualization of the underlying ontology helps to summarize information and to identify related context. This is explained in more depth and with some examples in the next section.

4.4.2 Visualizing the Fuzzy Grassroots Ontology

With the fuzzy grassroots ontology the semantic context of a query can be taken into account. In searching for a certain term, the fuzzy grassroots ontology enables communication operatives not only to find precisely the term but also (vaguely) related terms. For example, if the communication operatives googles the term Apple, all the results that include exactly this term

will appear on Google's SERP. Besides noise (e.g. mentions of the fruit and others; see also chap. 2) the communication operatives will find news describing the rivalry between Apple and Microsoft[28] (e.g. concerning innovative cloud computing services); internal or external information about Apple; the Apple store and a blog dedicated to Apple products.

According to [Portmann & Meier, 2010; Portmann, 2011a] all of these results were found because they contain the searched-for term (i.e. Boolean search). In contrast, the result of a fuzzy grassroots ontology-enhanced search considers also related terms, so-called suggestions. According to [Hearst, 2011] searchers often ask for a GUI that organizes search results into meaningful groups in order to better understand the results and to more naturally decide what to do next. Since the fuzzy grassroots ontology was created from folksonomies, it is likely to state that there are numerous apple-containing tags and that all underlying documents can boast also other tags like Microsoft, cloud computing, good or bad, and so on. Hence a fuzzy grassroots ontology-enhanced query comes in addition to a Boolean query along with other search results. This allows finding related topics and constitutes a big advantage for online reputation analysis. If Apple, Steve Jobs, or cloud computing are often tagged together with Microsoft or Bill Gates for example, then all these terms are also included in the results and therefore shown on the Topic Map. This is visualized in figure 4.7. The user can easily see that, and based on a fuzzy membership degree (i.e. u_{ij}; see chap. 3) to which degree a certain topic is related to Apple. The membership degree u_{ij} thereby is visualized by the length (i.e. depending on the distance d) of the edge of a tag to a topic or a topic to another topic.

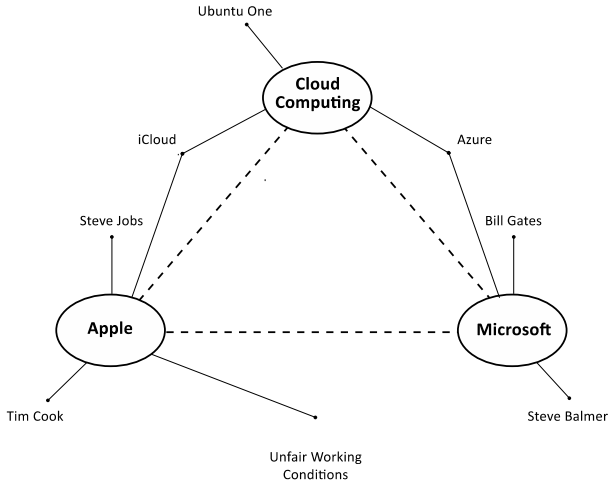

Figure 4.7: Topic Map Example.

Consider, as another example, that Leontien Aarnoudse's scathing post about Apple's unfair working conditions at a Chinese supplier of iPad parts has no wide coverage and his entry does not appear on the first few pages of

Google's SERP. Nevertheless, the readers of his blog can tag this article with the name of the organization (i.e. Apple) and some more attributes (e.g. unfair working conditions, etc.). Here, if a query for the organizations name is enhanced by fuzzy grassroots ontology also related tags appear on the Topic Map. The fuzzy grassroots ontology-powered Topic Map now can reveal these related tags too as they co-occur with the organization name. By clicking on the unfair working conditions tag, a concerned communication operative is now able to find entries about worker in China working for Apple in unfair working conditions as [Aarnoudse, 2011]'s covers in his blog entry. Emerging topics, in this case the scathing article, appear at the edge of the Topic Map behind other related topics (e.g. competitors, stores or products). Yet, they appear and can be spotted on the first day of publishing the blog entry.

The weak signal problem circumscribes the fact that it is often difficult to find new reputation issues as long as only a few unknown stakeholders (e.g. public groups) are involved [Weick & Sutcliffe, 2007]. The fuzzy grassroots ontology helps alleviate this problem. Even if only few social media users write about a certain topic, communication operatives can spot it on the Topic Map. This allows communication operatives to react faster and to intervene in a phase where the (traditional) media do not yet report about it. Yet, the more stakeholders are involved in an issue the more difficult and costly it gets to circumvent it. At the beginning the scope of action for an organization is much higher at much lower costs. Hence, thanks to the fuzzy grassroots ontology, potential negative effects can be lessened and new promising developments can be spotted before other stakeholders (e.g. competitors or media) become aware of them.

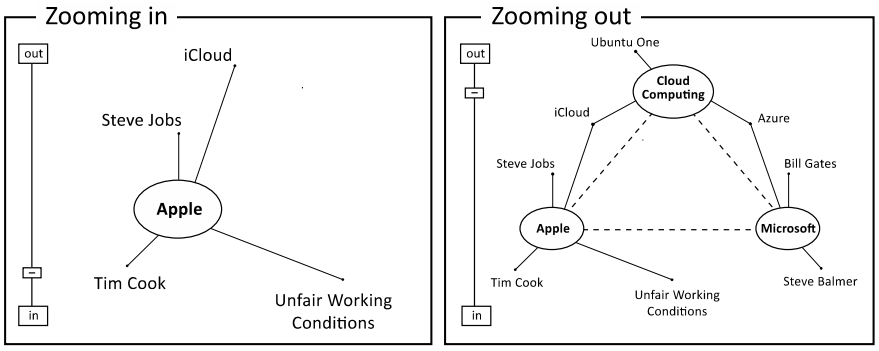

Figure 4.8: Topic Map Zoom Function.

It can be stated that with the fuzzy grassroots ontology scanning becomes much more efficient. Another advantage lies in the fuzzy grassroots ontology's zip. Environmental influences (e.g. trends, developments, opinions, etc.) are constantly taken into account. Depending on the constantly updated fuzzy grassroots ontology, the Topic Map is permanently adapting itself to new circumstances. Hence the interactive Topic Map does not only permit to

find emerging topics but also permits to overview changes in well-known topics (i.e. in the ensuing monitoring). With a zooming in-and-out function (see fig. 4.8), for example, a communication operative can define the scope of a query [Hearst, 2011]. The more restrictively the zoom function is held, the more a specific term has to be related to the searched-for term in order to appear on the Topic Map.

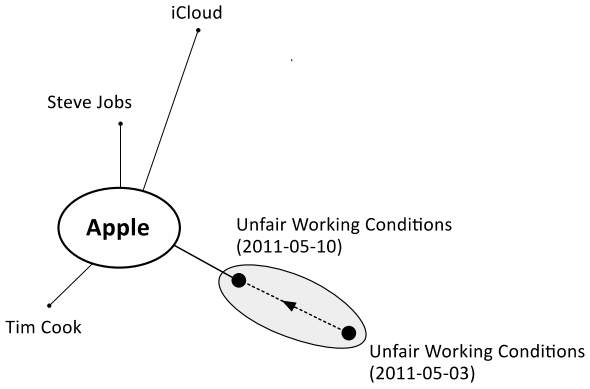

Figure 4.9: Monitoring a Topics Progression over Time.

If a topic is about to emerge it is only visible if the zoom is held open and imprecise. A later query may reveal that the same topic is now also visible if the zoom function is close and quite strict. That means that the relevance of this topic towards the organization has increased. This increased progression is shown in figure 4.9.

4.5 Further Readings

There are numerous books concerning reputation management. A quick differentiation can be made between more sociological or more business and management oriented literature: Readable from sociological oriented perspective are the books [Fombrun & van Riel, 2008] and [Peters, 2011]. From a business and management perspective, a read of the books from [Aula & Mantere, 2008; Doorley & Garcia, 2010; Griffin, 2009; Beal & Strauss, 2008] is highly recommended.

There is a lot written about offline reputation management, little about online reputation management, and almost nothing about how to integrate the two. A holistically approach, however, is critical because online and offline reputations easily flow between traditional and social media. Since today an organization cannot control what is said, it is better to interact in the Social Web. [Beal & Strauss, 2008; Eck, 2010; Peters, 2011] mainly go into online reputation management in the Web. In this manner they advise for a successful reputation management to become fundamentally transparent and as a result trustworthy. Thereby they also propose a holistic approach of online and offline reputation management.

For the technical aspects of reputation management, especially the topics of issue management are important to consider. [Robert, 2011; Ingenhoff, 2004; Röttger, 2001] showcases in their books the topic to its full extends. Thereby these books also cover the handling of crisis, an important task if stakeholders trust is lost in one fell swoop, for example. [Thiessen, 2011] presents to this end literature particularly concerned with crisis management. However, [Sterne, 2011] illustrates metrics to measure all these online reputation management efforts.

The creation and the use of the fuzzy grassroots ontology is subject of [Portmann & Meier, 2010; Portmann, 2011a]. Besides, [Portmann & Kuhn, 2010] presents cartographic visualization as another example of knowledge representation. Topic Maps are explained by [Pepper, 2010]. [Kosko, 1986] for the first time presents fuzzy cognitive maps, namely for representing casual reasoning as fuzzy-graph structures. Nearly a quarter century after Koskos publication, [Glykas, 2010] presents research efforts in the development of fuzzy cognitive maps. Last, the use of the fuzzy grassroots ontology for online reputation analysis can be found by [Portmann et al., 2012].

5 Requirements for Online Reputation Analysis

"Everything is vague to a degree you do not realize till you have tried to make it precise."

—Bertrand Russell

Responsible communication operatives feed social media elements not only daily but hourly these days. The intention is to build trust through quick, truthful, and transparent communication [Beal & Strauss, 2008]. Moreover, traditional press releases are discussed in blogs, microblogs, social networks, and sometimes even in wikis. Products are assessed through rating systems; advertisements are displayed, tagged, and commented: but above all, crisis situations are taken to external social media elements, where the consumers, mostly outside the reach of the organizations, discuss them. To this end, [Thiessen, 2011] provides a conceptual framework for communication in crisis situations. In the course of this an organization's quick, truthful, and transparent communication turns out to be a key instrument for sustaining reputation during a crisis.

Since Web search engines prevalently score social media elements higher than traditional press releases, these elements are listed at the beginning of SERPs. Thus, their retrieval is not only easy for consumers but also for journalists and competitors. Notwithstanding an organization's best effort to build a positive reputation, there will be circumstances when the organization's reputation faces an assault from social media elements. Unfortunately it is quite difficult to predict or prevent such reputation assaults. As presented, target-oriented online reputation analysis can provide, at least partially, a remedy. Based on such an analysis, an organization's communication operatives can detect early mentions about the organizations name, brands, service, products and executives early, even if nothing is explicitly mentioned about it [Portmann et al., 2012]. Whether or not this is the case, this

causes the ultimate chance to show the stakeholders that the organization is listening to them and learning from their criticism.

However, to approach fuzzy online reputation analysis, its various needs must be investigated. Thereby these needs are documentations of what the FORA framework should be. To methodologically understand the different needs, this chapter is split into five parts: In section 5.1, an example scenario is introduced. This scenario is rooted in a real event, which is extended based on experiential knowledge and comprises a thought experiment to give the reader an idea of the issues of online reputation analysis in the Social Semantic Web. To cut requirements for the FORA framework and to emphasize the scenarios significance, section 5.2 presents three case studies: The first illustrates the use of social media elements in organizations, the second the challenges of journalists relying on Web search engines and the third the challenges of communication operatives performing online reputation analysis. Deduced from the scenario and the subsequent case studies, section 5.3 presents the communication operatives' requirements for online reputation analysis. Thereby implications for the FORA framework are drawn. In section 5.4, the communication operatives' requirements are summarized to technical implications for the FORA framework. Last but not least, section 5.5 completes this chapter with suggestions for further readings.

5.1 THE APPLE INC. SCENARIO

To give an impression of the issues of online reputation analysis in the Social Semantic Web, a small scenario is presented. This scenario is rooted in a real event at Apple Inc. and extended with a fictive outlook. When Apple introduced the long-awaited iPhone on Friday June 29, 2007, the product was a triumph by any standard [Beal & Strauss, 2008]. After months of waiting for the organization's initial advances into smartphones, consumer queued in lines on the street—each of them keen to be one of the first to pick up one. Loyal customers as well as converts bought in the first two days of purchasability about a quarter million iPhones. With the most popular model selling for $599, Apple estimated to add millions of dollars to its bottom line and envisioned within two months a stock price increase from $129 per share to $144. Apple's reputation as an organization that produced high-quality, innovative products that allures customers was firmly strengthened. The organization was on a winning streak, and there was ostensibly nothing that could stop its triumph.

However, on Wednesday morning, September 5, 2007, suddenly Apple's former CEO Steve Jobs made the announcement that it was reducing the iPhone's price from $599 to $399. The decision came as the organization struggled to adapt the iPhone's price with the iPod music player, while making the smartphone more appealing priced for the imminent holiday shopping season. Steve Jobs backed up the decision by saying: *"It's very clear we have a breakthrough product on our hands, but it's also clear that many can afford it, some can't. We'd like to make it affordable to even more folks going into this holiday season."* [Beal & Strauss, 2008].

This information sent shockwaves throughout Apple's stakeholder communities. The traditional (mass) media wondered whether the iPhone sales were slacking off while investors were flabbergasted by the sharpness of the price reduction effectuating the stock price to drop five percent by the end of this day. In the meantime, customers (i.e. evangelists willing to stand in line to be among the early adopters) felt they had been tricked into overpaying their iPhone. Journalists covered the organization with questions, bloggers discussed whether customers had been fooled, and social networks were buzzing with critical comments [Beal & Strauss, 2008]. Apple's former immaculate reputation was severely damaged, and the organization needed to move in order to avoid leaving a bad taste in the mouths of its stakeholders.

Apple's iPhone faux pas is an ideal example scenario of just how promptly a crisis can escalate on the Web. Although the organization has one of the world's most cherished brands, it simply did not foresee the large-scale counter-reactions induced by the price reduction. While it may look like the organization had tripped up for the first time, it merely faced its first rebellion of the social media incident. The iPhone failure happened during the growing adoption of blogs, social networks, and other social media elements

as a platform for sharing complaints. Customers had criticized Apple's brands, service, products and executives before, only now the expansion of social media ensured that Steve Jobs, Apple, and a few million others online, heard the message in a clearer manner.

Consequently, since then, and in order to prevent the repetition of such a fiasco, Apple pushes online reputation analysis ahead. Hence Apple's CEO Tim Cook wants to know all mentions about Apple's brands, service, products and executives. Since sometimes discussions in the Social Web first manifest quietly, then surge and finally pelt on an organization with full intensity, he wants furthermore also to know if and how fast a topic is gaining or losing in importance over a time period (see fig. 4.9). At the moment such a problem, for example, could pose the raised unfair working conditions at a Chinese supplier of iPad parts (see chap. 4) [Aarnoudse, 2011]. Hence, to spot early dissonance, a sound online reputation analysis should include fuzzy indications not directly mentioned by the underlying social media elements. In addition, during the analysis, issues should be summarized into related topics. For example, this can be done using an application that is based on a constantly updated ontology and its interactive visualization as Topic Maps.

Imagine now that in order to implement a continuous online reputation analysis, Apple's CWO Philip Schiller and his corresponding communication operatives field an adequate application that supports them in the endeavors of online reputation analysis. For that purpose they make use of an online reputation analysis application. This application is based on the up wind Semantic Web technology, because the Semantic Web makes a pledge to enable coming computers to understand the semantics of Web contents. New developments (e.g. layers) are built on other layers that are already implemented. Thereby these new layers allow the best possible further development of the Semantic Web. In this fictive scenario, Apple's application thus avails from the Semantic Web model of ontology to summarize issues into related topics. However, an instability factor for the application is the prosumers vagueness in their use of language (i.e. semantics). This vagueness includes imprecise concepts like beautiful, lightweight, and large (see chap. 3). To overcome this instability factor, the application draws on fuzzy logic, the most promising technique for dealing with semantic vagueness.

The problem with new and previously unobserved information on the Web is that the relationship between terms and topics is not precisely known. Now, to scan mentions in relation to the organization, Philip Schiller, in this scenario, first enters the search term Apple in the search field box on the start page of the application's dashboard. On the left side of the dashboard, immediately, a fuzzy grassroots ontology-based knowledge representation (i.e. Topic Maps visualizing all Apple-related topics and terms; see chap. 4) appears. The fuzzy grassroots ontology constitutes the attempt to capture the vagueness of human concepts. The visualized return as Topic Maps

should be considered as fuzzy approximation. Moreover, on the right side of the dashboard, a hit list appears. This hit list, in reality, is not conventional; it is partitioned into context dimensions. These context dimensions come across as any information (no matter how fuzzy) that characterizes a situation related to the interaction of communication operatives with the application and the according social media elements (i.e. stored in the application's underlying knowledge base). Accordingly a desirable distinction of context would let him answer the questions of who said what, when, and where—the minimal categories for perceiving context [Dey & Abowd, 2000; Hohenberg, 1978; Dey et al., 2001]. Yet, social media elements hit lists are portioned according to these minimal context categories. On the dashboard they are visualized together with the ontology-powered Topic Maps. Figure 5.1 illustrates the dashboard of Apple's online reputation analysis application.

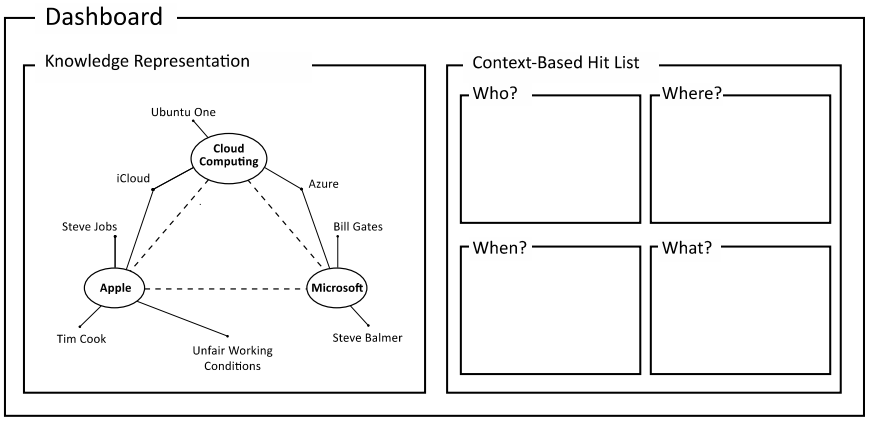

Figure 5.1: Online Reputation Analysis Application Dashboard.

Hence, next to the interactive visualization of the ontology on the dashboard (i.e. using Topic Maps) also the social media elements hits are interactively visualized. In this way it is possible to browse the ontology and simultaneously discover underlying social media elements sorted according the context dimensions.

Table 5.1: Social Media Elements in Context.

		Context dimensions			
		What?	When?	Where?	Who?
Social Media	Weblogs	✓✓✓	✓✓	✓	✓
	Microblogs	✓	✓✓✓	✓	✓
	Folksonomies	✓✓	✓	✓	✓
	Social Networks	✓	✓	✓✓	✓✓✓
	Wikis	✓✓✓	✓	✓	✓✓

Table 5.1 illustrates potential social media elements discrimination criteria as input sources for the four dimensions of minimal context. To illustrate, the hits for the context dimension what are coming from Delicious service

(see chap. 2) because Delicious includes mostly descriptive information. However, for the context dimension when, the hits are coming from Twitter service (see also chap. 2) because Twitter provides additional data such as the date and time of a tweet. Thus, it is possible to find past as well as real-time information depending on the Web agents' frequency of updating the crawl frontier list (i.e. a list of URIs for the Web agents to visit) and, in doing so, also the knowledge database. In table 5.1 the respective allocation of social media elements with appropriate context dimensions was created on a three-stage scale as a suggestion. Note that the membership of the respective social media element to a particular context dimension often cannot be determined precisely but fuzzy. In addition, UGC is never precise because online data are in a constant flow and change continuously.

Nevertheless, the consequence for Philip Schiller and his team is that they not only find more relevant information concerning their entered search term Apple but also receive more structured information (based on the ontologies and the context dimension partitions). Their application queries several top social media search engines, combines the results and generates a hit list. With a conventional Boolean search system, they would only find information containing the term Apple. In contrast, the innovative application enables, to find not only the search term but also more or less related topics and terms. Since the dashboard is interactive, Apple's communication operatives can zoom in-and-out and with the help of Topic Maps, browse the ontology in a straightforward manner. Based on the ontology, a search for Apple can also yield accordingly labeled hits concerning the business competitors Microsoft and its related terms (see Ch. 4). By availing a click-function on the hits to visit, the responsible communication operatives now can directly interact with a specific underlying social media element (e.g. a weblog or a microblog entry).

Because the application's found social media mentions are supported through an accentuation via colors, they can smartly respond to hotspots. Negative social media mentions are presented in the hit lists marked in red, positive mentions green and neutral mentions yellow. To do this, the application applies NLP to identify subjective information in the source material.

A further possibility of the application is to store once found reputation issues. Thereby not only the issues can be stored for later processing but also the whole query itself. So, Philip Schiller is able to store the Apple query from above once and invoke this query when necessary at all times, for example. Another important point of his application is that Philip Schiller and his team can start notification queries. Based on a trigger function, the application finds every mention containing the queries search term and notifies him about this mention. On the basis of this notification they can now extend their search query and interact with the dashboard to eventually find additional Social Web contents. So it is possible, for instance, to set an iPhone

trigger and to get a notification every time iPhone is mentioned somewhere in the Social Web. This is comparable with Google's Web content change detection and notification service. In contrast to Google Alerts[29], however, also a search range adjustment is possible. For example not only the exact term iPhone but also related terms (i.e. terms as iPad, smartphones, etc.; see chap. 4) can in this way act as a trigger.

Last, for reporting reasons, Apple's communication operatives can now download the found issues as list (e.g. as a spreadsheet). This spreadsheet helps to communicate with the CEO and the management in general. Moreover, the downloaded spreadsheet can be loaded again into Apple's reporting system in an easy way. Such reporting is a fundamental part of larger movements towards improved business intelligence and knowledge management. Because of the reporting, and in the sense of an integrated reputation management, concerned departments can counteract. Afterwards the effects can be controlled by specific social media measurements (see chap. 4) [Sterne, 2011].

5.2 Case Studies

The Apple Inc. scenario highlighted an anticipated online reputation analysis application. The question now is to see if such an application is correspondingly favored by various media users. Evidence from multiple use case studies for or against such an application, and to more profoundly and scientifically investigate online reputation analysis, should be considered [van Aken, 2004].

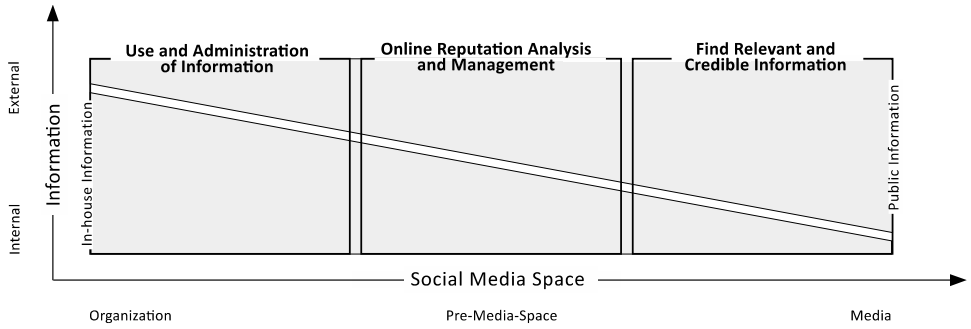

Figure 5.2: Case Studies Continuum.

Within this PhD project, three case studies were conducted according a social media continuum, ranging from an organization via the pre-media space through to (traditional) media (see fig. 5.2). That way a wide range of online reputation analysis is covered. The case studies indicate the collection and presentation of detailed information about average media users, including an organization's perspective whether to participate in social media elements, a media perspective about the Social Semantic Web as research tool and lastly a perspective of the analysis of the pre-media space. The infor-

mation power in figure 5.2 decreases with a shift from organization-owned to media-owned (i.e. grayed out).

Each of these cases was drawn up in collaboration with a responsible post-graduate student at the University of Fribourg (Switzerland)[30]. Therefore in the following sections these students' Master or Licentiate theses are quoted. However, their theses originated in a close cooperation. Section 5.2.1 illustrates the first case study: Mainly based on an in-depth literature review, [Fuchs, 2010] elaborated in his Licentiate thesis in a narrative way, why it is worthwhile for the CD Lab Inc. (a provider of software for sewer inspections) to use social media elements. In section 5.2.2, the second case illustrates the challenges of journalists relying on Web search engines: In her Master thesis, [Hächler, 2010] elaborated with qualitative interviews the potential of the Web as a research tool that meets professional journalistic standards. Afterwards, in section 5.2.3, [Uhlmann, 2011] clarifies in the third case the challenges of online reputation analysis instantiations and issued in her Master thesis a set of requirements towards online reputation analysis applications. Also with qualitative interviews but followed by test installations, this case classifies and categorizes online reputation analysis in Swiss financial services institutions.

As a form of qualitative descriptive research, case studies look intensely at individuals or small participant pools, drawing conclusions only about that participant or group and only in that specific context. Hence, these case studies are by no means intended to focus on the discovery of a universal, generalizable truth, nor on the look for cause-effect relationships; instead, the emphasis is placed on exploration and description. Moreover, the illustrated boundaries (see fig. 5.2) are not that absolute as they appear but rather fluent. As a consequence the information power shifts but is never hard.

5.2.1 Cooperate rather than Coordinate

The first perspective considers social media elements in general from an organization's standpoint. As a first step to unveil the potential of social media elements' commercial use, an analysis and evaluation in organizations environment was conducted. For those reasons a wide-ranging literature review concerning social media elements and its usage in organizations was undertaken. In a second step, the findings were pieced together. Following [Flyvbjerg, 2006], which recommends an active participation within the object of study, as last step in the teamwork with Roger Fuchs a promising social media application for knowledge management was implemented within the CD Lab Inc.; picked as a representative organization of a typical Small and Medium-sized Enterprise (SME) in Switzerland. Throughout the analysis of the case, a wiki as knowledge base for supporting activities was created. In doing so, the potential of social media elements in general and wikis in special was experienced firsthand. In this PhD thesis first and foremost the changes of organizations in relation to social media elements are of interest.

Such being the case, at this point only these findings are briefly presented. The implementation of the wiki is therefore neglected. However, for an interested reader [Fuchs, 2010]'s well-written Licentiate thesis is highly recommended.

The findings underline that the Social Web's triumph in private context cannot necessarily be extended to an organization's context. The Social Web contains, as already explained, socio-technological elements in which the human factor plays an important role. However, since an organization's environment differs from a private environment, difficulties and challenges arise that in non-commercial use bear little effect. Hence, the differences in the usage of technological communication, coordination and cooperation instruments within and outside an organization have never been as big as in the Social Web era [Hein, 2009]. Since the Web has socialized and has become the Social Web, the number of possible communication channels multiplied. Thereby social media elements emerged that profoundly changed and enriched the communication and organizational culture of everyday life. In the twinkling of an eye the organizations had to face a number of potential communication channels and collaboration implementations that are versatile, and more importantly allow their systems to partly overlap each other. Both for internal as well as for external communications these socio-technical developments have implications.

As long ago as 2006, in every second organization more than three-quarters of employees [Hein, 2006] predominantly worked with knowledge (i.e. the employees performed tasks of intellectual nature). This percentage has unquestionably increased in the meantime. Likewise the routine of certain social media elements was in the intervening time at a great pace integrated into everyday life and work processes too, so that there is no longer a clear personal and professional partitioning. The increasing impact of social media as a communication channel between external and internal stakeholders and the organization itself boosted the trend towards an integrated work life model that makes it hard to control the whole employee's work environment [Hein, 2009]. However, a fully controlled work environment can entail a number of negative implications. For example, the self-selection of employees may be interfered, the collaboration with suppliers may suffer, and most important, the reputation in general can be severely harmed.

It is assumed that ultimately the organizations that have the biggest success perceive social media elements as cultural interfaces and not only as mere communications channel [Hein, 2009]. However, if an organization only depends on traditional media elements, its scope is limited and its communication stays asymmetric (i.e. top-down) [Portmann & Hutter, 2011]. An organization's complete control of all channels and content is hardly possible. When needed, employees sidestep on public social media that provide the desired communication elements. After all, who became acquainted with blogs as a means of an individual and open exchange of views or with wikis

for collective work on documents and projects, will bring control mechanisms and institutional editorial into questions and if necessary switch to external offers. Using social media elements such asymmetries can be reduced to a minimum. In addition, social media elements encourage processes of internal integration as well as external adaptation [Hein, 2009].

As elaborated by [Fuchs, 2010], social media elements afford immense potentials for private as well as commercial use. The crucial point is on one hand the ability to distinguish the right element for specific needs and, on the other hand, the competence to use this element effectively and efficiently, since with the multiplicity of systems, applications and services the options grow. A media convergence can be observed in both, technical and structural realms. This could possibly lead to a reduced complexity, but on the other hand it is assumed that in the course of data aggregation and further interconnectedness, in turn, frequently new elements will be provided. As a result, media competence will play an increasingly important role in the future [Fuchs, 2010].

Regarding the use of social media elements in business environment, it is assumed that once a critical mass of organizations is availing these elements, also less innovative organizations will consider using them. Yet, according to [Fuchs, 2010] many SMEs in Switzerland classify today's free and open-source social media elements as poor-quality and second-class. If social media elements work out to gain a reliable status, then their image will alter and they will give proof of their possibilities as corporate social software.

With the Social Web not only the WWW has changed. The increasing interconnectedness and the penetration of all spheres cause that communication, collaboration and administration to have changed significantly too. The bottom-up mentality of the new Web era and the relative uniformity of information access led the Web to become social. At least on a virtual level social distinctions blur, and constitute a real democratic participatory convergence culture [Jenkins, 2008]. Hence, organizations are becoming aware that the Social Web is ever more important to their business. Thus, these organizations try to include these social media elements into their regular activities.

To interact where necessary, more and more organizations rely on scanning and monitoring reputation issues on the Social Web. These at the moment rather large organizations, which do so, can have a leading role for SMEs. Consequently, an online reputation analysis application as presented in the scenario could support the organizations' communication operatives. As presented, for a modern and effective online reputation management, rules must be modified and extended [Peters, 2011]. This obliges a synchronization of individual responsibilities of online reputation management in the corresponding departments as well as within corporate communication itself. On the other hand, existing rules, for example in dealing with customers or journalists, should be completed by Social Web guidelines. Based on these

modified and expanded rules and processes, a strategy for online reputation management can be developed. This strategy begins with the planning of activities to build, maintain and expand an organization's reputation.

The next section takes a closer look at (online) journalists and their research methods to get relevant information. Online journalists are located at the opposite direction of the continuum. In particular these journalists use the organization-provided information as long as they can reach it.

5.2.2 The Social Web as Research Tool

This second perspective considers traditional (mass) media. It has been investigated together with Livia Hächler, a postgraduate media and communication Master student. [Hächler, 2010]'s interests were in the challenges journalists face when dealing with Web search engines. In her Master thesis, she primarily elaborated the potential of the Web for research issues with qualitative interviews among online journalists in Swiss online news outlets. As a point of reference, the momentary situation of the consumption of the Web for research purposes has been surveyed. Besides, expected requirements for future Web search engines have been ascertained. In the first place within this PhD thesis the changes for online journalists concerning social media elements are of importance. Such being the case, at this point only these findings of Hächler are briefly presented. However, for an interested reader [Hächler, 2010]'s detailed survey is very much endorsed.

To an online journalist, the most significant and far-reaching ramification of Web conversations becomes apparent by the acceleration of the news production. As medium hurdling barriers of time and space, the Web is a mixed blessing: For one thing it boosts fast exchange of information, as a result making newsgathering simpler, even overcoming geographical boundaries. For another thing it erodes at the same time the established basis of existence, leading to sparse resources and radically increased time pressure. Especially online journalists are faced with shattering time constraints for two different reasons: Firstly, online news publishing is a matter of minutes. Secondly, the Web has changed the readers' behavior causing a 24/7 news cycle that is steadily consumed in snatches. Achieving the professional designation regarding uncovering the truth and research requires time, and time has become very limited in an online environment. As a result, the Web has become to be an important tool to do research quickly, simply because it is fast.

This is where the interest to analyze today's use of the Web as a research tool arose out of. The main goal of this research carried out in the Swiss online news outlets was to identify the current situation in online editorial offices concerning the use of the Web for research matters. Thereby journalists also include information from organizations (see sect. 5.2.1). Nevertheless, the ulterior reason was to harvest as much information as possible to be able to make a picture about the current use of the Web as a research

tool. To get this goal, [Hächler, 2010] has chosen qualitative interviews as her method. Hence, the samples and the resulting information cannot be seen as representative of the whole journalistic field, but as an indicator of the general accepted and heterogeneous ways in which online journalists rely on the Web for their research [Hächler, 2010].

Generally [Hächler, 2010] found out that the way the different journalists use the Web as a research tool differs strongly. Various reasons make for this situation: Firstly, depending on the journalistic field, the different tools can alternate. For boulevard journalism, social networks (e.g. Facebook or Twitter) seem to be of great value. In contrast, a news agency is much more conservative, buttoned-down, and cautious in using Web items because of the eventuality of it having been gerrymandered. Apart from that, journalists in specialist fields do find experts via the Web (e.g. via blogs or microblogs) and enjoy being able to read (international) news as an additional input. The only tools used in a similar way by all interviewees seemed to be Google and Wikipedia. Those are used as a starting point and as an access to research, which help attain a general synopsis of the issue. What differs once again is the way the information found is incorporated into the story to be written. Where to some journalists Wikipedia, for example, is simply too unsure and too susceptible to manipulation, others do integrate such data straightforward by doing a reasonableness check. This is especially the case when the information is considered as very factual and insensitive. A double-check is then omitted mainly due to time constraints.

Time and time pressure seem to have a large influence and are an important factor in the use of online sources. For one thing, it impinges on the choice on accounted reliability and credibility of a source. Many social media elements are viewed as precariously; and found information would need to be verified. Often the time for such validation is lacking, what results in either abandoning the source or using it by solely verifying its plausibility. According to [Portmann & Hutter, 2011; Hächler, 2010], this can be observed very well when a (catastrophic) incident has happened and pictures or statements are quickly being sent via social media elements (e.g. Twitter or Google+; see chap. 2). However, media workers are becoming evermore inclined to see social media elements as a valuable adjunct to traditional media.

For another thing, it affects the journalist's opportunity of even becoming acquainted with the new tools. Especially tagging was found to be obscure among media workers, basically because of the lack of time and incentive to even experience it. Those that know of it do not seem to consider it as relevant. Yet, most journalists have only heard of it or do not know of it at all. Social bookmarking was not known as a choice of coming up with interesting online sources. In some cases the same could be detected with blogs and microblogs. These tools are time-devouring if one really wants to become acquainted with them and find authors of high quality. In the journalistic

day-to-day life it is often not possible to find such an amount of time. However, following [Portmann & Hutter, 2011], many journalists who were dismissive about it have changed their perspective in the past few months as the value became apparent, for example in the coverage of the Arab uprising, the Japanese earthquake, and the occupy Wall Street movements.

Nevertheless, time alone is not the only reason why social media elements are not being trawled. The Web embodies an immense amount of information, mostly considered irrelevant to a journalist's business. By providing more raw material than ever from which to distil the news, social media have both done away with editors and shown up the need for them. Though the principle is the same in real life—everyone can give journalists inputs, hints and background information, but in journalistic work it is crucial to ask the right person. Any online or offline information needs to arise from a relevant and eminently credible and reliable source. Since on the Web much is vulnerable to tampering, finding a trustworthy source seems to be difficult. In order to know what can be considered as trustworthy, journalists seek to know the sender of the information. For example, official governmental sites are then treated as more reliable sources. The trustworthiness of private company websites hangs on their generally conceived credibility. Data might be taken, always quoting the source and referring it as an aspect of the story. This transparency seems to signify a new kind of journalistic objectivity. Time pressure usually prevents media workers from doing all the verification, which theoretically should be made before using a source. Down the road, the general journalistic principle of having more than one reliable source and verifying its reliability by way of experience and also direct (telephone) contact, leads all journalists to carry on online research as much as it does offline research [Hächler, 2010]. Commonly the Web is used as an additional research tool, particularly for background information. Certainly, this medium has accelerated traditional ways of information gathering at a high rate too. The examples here range from finding online studies and press releases, to having live streams (e.g. of press conferences) through to finding telephone numbers of the contact sought-after.

When talking about favored prospective evolutions, many inputs related to better and more specific Web search engines were stated. These wishes either included Web search engines that increasingly connect various sources and databases to lessen the need to search through many different portals, or involved being able to do a more unique search, which would lead to only non-copyrighted pictures, qualitative local searches or specific branches. In general, the interviewees mentioned finding relevant results better and quicker, especially when only having few keywords, would be of great help. Google's lists seem to be easy to handle, but much information is missed ultimately because it was maybe not listed on the first few pages of a SERP.

Different results-visualizations could help in tackling these issues. However, most journalists do not like the idea of giving computers the power of mak-

ing decisions on relevance and importance of information. Even though journalists, as all Web users, need technological help in keeping control over the vastness of information and finding the relevant drops, they fear of losing control in deciding where to search and what is relevant.

Taking a deeper look into the nature of the Social Web and its tools as well as that of the Semantic Web, it can be concluded that the central problem in employing online sources is their trustworthiness. The Web's openness and participative nature, is at the same time its drawback because the information is never completely trustworthy in regard to modern ways of ascribing trust. One way onward is to quit the ideology of viewlessness and take on board that (online) journalists have a range of views; to be open about them while holding the reporters to a basic standard of accuracy, fairness and intellectual honesty; and to use transparency, rather than objectivity, as the new foundation on which to build trust with the audience. This transparency also means linking to sources and data, something the Web makes easy. It is, as a communication platform, flattening hierarchies and molding public communication in prior unknown ways. To use the Web to its fullest, [Hächler, 2010] concludes that these developments might entail some fundamental paradigm shifts in society as well as in the field of journalism. However, online journalists can find reputation issues during their searches. If there are several sources pointing in the same direction, the chances are high that a (online) journalist also uses only partially verified sources. Consequently, from an organizations standpoint, it should be of utmost interest to manage appropriate an organizations online reputation. Apple's online reputation analysis application presented in the scenario can help a responsible communication operative to detect such reputation threads.

The next section highlights the intermediary layer between the organizations and the media. This intermediary layer (i.e. pre-media space) is considered from a communication operative's viewpoint, which mediates between internal and external information of an organization.

5.2.3 Online Reputation Analysis Test Bench

The last perspective of this section considers online reputation analysis of the pre-media space. This comes across as the area in-between the two previous cases. Yet, this area is not controlled entirely by any of the aforementioned Social Web participants; rather it is an integral component of both. Predominantly large organizations perform online reputation analysis, as [Fuchs, 2010] discovered, but in future these organizations may have an impact on SMEs. This leading role is the reason why financial services institutions are chosen as representatives for large Swiss organization.

Katrin Uhlmann works besides her Master studies part-time at the communication division of PostFinance Inc., which is the fifth largest financial services institution in Switzerland. In collaboration, a set of requirements for online reputation analysis within these institutions was developed.

[Uhlmann, 2011]'s interest thereby was in the classification of online reputation analysis applications. In her Master thesis, she first analyzed the state of the art of online reputation analysis applications and ensuing compiled a set of requirements to these applications. Within this thesis the main interest is in the communication operatives requirements toward a most helpful application. Furthermore, the arisen set of requirement, in the end, is used as evaluation base. Nevertheless, also [Uhlmann, 2011]'s Master thesis is entrusted as a good read to interested reader.

The global financial crisis has severely deteriorated the reputation of financial services institutions and provides them with new challenges. Thought, this offers the institutions the unique opportunity to reposition and to bring themself into a solid constitution for the future [Heinrich et al., 2010]. According to [Kunert et al., 2011] maintain nearly two-thirds of the largest Swiss organizations their Social Web presence, but not that many of the financial services institutions. Those who do not maintain their presence, in particular are afraid to be fooled by hype, and therefore shy away from corresponding investments. Nevertheless, [Uhlmann, 2011]'s survey illustrates that already a few institutions operate their own Social Web presence after all. However, here too, the influence on the Social Web is rising. These days there are a number of online reputation analysis applications at hand for a systematic scanning, monitoring and controlling of social media elements. These applications provide options for quantitative and qualitative analysis, as well as for minimal semantic analysis. By most of these applications fervently debated issues can be identified, stakeholders detected, and topics analyzed. However, since such applications do not work fully automated, manual effort by responsible communication operatives is still required.

To evaluate the requirements of the communication operatives who deal with online reputation analysis, semi-structured interviews were accomplished in six Swiss financial services institutions. These interviews yielded a list of requirements for online reputation analysis applications. Afterwards, using test cases, three online reputation analysis applications of different origins and price range were explored at PostFinance. Based on test installations, it was evaluated to what extent each application meets the named requirements. The evaluation illustrated the current limitations of a wide range of applications. The wish of the interviewed communication operatives for an automated application, ready-made reports, and a minimization of manual effort remains generally unfulfilled. In the tested applications, the post-processing of the results through a communications operative is still very important. With a high number of hits, as it is the case in large organizations, this portends extra effort [Uhlmann, 2011].

The online reputation analysis' aim is to recognize reputation-related issues as early as possible, allowing organizations a better scope of action and control over their reputation. It is about the search for occurrences and trends that could emerge and affect the organization. Online reputation analysis

constitutes more than just a form of Web monitoring but rather a continuation of the strategic communication [Fuhrmann & Wewezow, 2010]. The qualitative interviews as well as the evaluation have shown that online reputation analysis applications turn out to be supportive and practical. However, the applications are so far limited by their characteristics. They observe pre-determined social media elements, search for postings therein, find and count them, and then evaluate the counted findings independently. Hence, mostly the applications are in their character oriented towards classical print media monitoring. Considering the immense mass of opinions in the Social Web this is hardly possible without significant manual effort. The present way to exclusively field these applications to define an organization's reputation is questionable because they comment only on limited stakeholders and, as a result, model no causal relationships. In addition, they mainly draw on the monitoring of communication flows, and in the course of this, especially in issues management [Ingenhoff, 2004], as well as crisis communication [Thiessen, 2011]. Online reputation analysis signifies a profound representation of global opinion about an organization on the Web, and provides detailed positioning analysis.

In summary, it can be stated that for online reputation analysis there is at the moment a broad uncertainty [Uhlmann, 2011]. While all interviewees discern the rising importance of online reputation analysis, however, there is no empirically established figure that is declaratory of what application is the best value for money ratio. It seems that up to this point, an unblemished application is non-existent. None of the evaluated applications meet the requirements of communications operatives at Swiss financial services institutions completely. Furthermore, there is uncertainty on what the specific area of responsibility of online reputation analysis constitutes, and how much effort should be invested. Therefore, recommendations can be issued only on an individual basis. A very important factor, if not the most important, is how much effort and budget is available for online reputation analysis.

5.3 Online Reputation Analysis Requirements

Derived from the previous case studies, this section presents the concentrated requirements to a fuzzy online reputation analysis application. Since online reputation analysis is a vague task that cannot be implemented with absolute accuracy, fuzziness as bridging link between humans and computers that support analysis fits well.

However, the mentioned requirements are an important effort into the verification process, since later tests should allow to be traced back to specific requirements. The presented specific requirements show what elements and functions are necessary for the application to automatically support all of the tasks that are common for identifying online reputation issues. All these requirements are always compiled with the sociological perspective of an organization's communication operative in mind. Yet, the perspective of

communication operatives are tailored to the Social Web and therefore sometimes complemented with requirements towards a socio-semantic information system. At that, socio-semantic information systems are information systems of the Social Semantic Web.

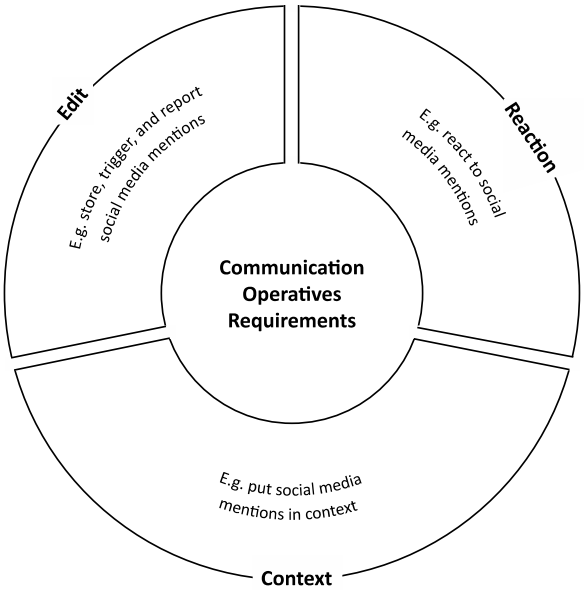

Figure 5.3: Online Reputation Analysis Requirements.

Together with Katrin Uhlmann the requirements of communication operatives in Social Web context concerned with online reputation analysis have been defined. Thus, these concentrated requirements stem likewise from the Master thesis of [Uhlmann, 2011]. Figure 5.3 illustrates the most important sociological online reputation analysis requirements.

- *Reaction*: For a communication operative it is important to be able to react to social media elements.
- *Context*: To do this, he wishes to put the located mentions into context. Only in this way he can grasp full coverage of a potential risky blog entry or a puff piece, for example.
- *Edit*: The last key point is to edit found social media mentions.

In the following sections, first section 5.3.1 highlights the most considerable requirement of a communication operative concerned with online reputation analysis—the ability to react to social media mentions in appropriate manner. In section 5.3.2 the found mentions are put in context to support the communication operatives in the searching process. For example, ontologies can help broadening the operatives' horizon and an automatic sentiment analysis can support spotting upcoming threads earlier. Finally, sec-

tion 5.3.3 is concerned with the also important requirement of communication operatives—the storing, triggering and reporting of found issues. To be able to react and put mentions in context iteratively (e.g. in an ensuing monitoring), it is important to trigger and store searches. Then, the found issues must be reported in an adequate format. Of course, these requirements can be extended limitlessly, so the presented requirements are the most important for a communication operative concerned with online requirement analysis.

5.3.1 React to Mentions

The social media elements can, as highlighted, enlarge an organization's marketing reach and thereby strengthen its consumer' relationships by engendering a dialog between them. However, the perception and use of social media is not consistent. Many organizations that could profit from the possibilities of social media have not integrated it into their digital ecosystem, while others are pursuing social media initiatives because they feel pressured to do so. They observe the benefits of social media, but they have no adequate application from which to make development decisions.

The most crucial point for a communication operative engaged in social media communications is, at least, to react to mentions about the organization. Since the Social Web is tremendous, consequently for the communication operatives often this presents insurmountable challenges. An application to systematically analyze social media mentions would be of tremendous help. Otherwise the Social Web's information overload feels like looking for a needle in a hay stack. For a communication operative, it is crucial to know how to extract relevant information. This leads to the question how such relevant information can be detected. Currently the problem with representing social media mentions and queries through sets of keywords is that they yield results which are only partially relevant to their actual semantic contents. As a consequence, the matching of underlying social media mentions to the query terms is vague. An application to support a communication operative's work is certainly committed to address these issues.

The goal of such an application is to make it easier for communication operatives to collect, organize, find, visualize, and share their information. Based on such an application an appropriate dialog with consumers could be formed. However, according to the online reputation analysis process (see chap. 4) the mentions need to be detected and prioritized. This detection should also include not directly named issues. Normally in a flame, for example, not each single organization is mentioned but the whole sector. In other words, a flaming against a whole industry sector should be detected too. Or an organization's disgruntled customer (or even employee) let off steam without directly naming the concerned product itself. Hence, also this should inevitably be detected by the application. The next section reveals a possible solution.

5.3.2 Put Mentions in Context

To support a communication operative in the quest for social media mentions of the organization, a corresponding application should broaden its horizon. Since not all mentions are named specifically, the application should help to recognize all these mentions and correlations in general. These correlations can range from simply revealing associated words mentioned often together to ontologies that, for instance, illustrate correlations on a grand scale. Rather than focusing just on SERP ranking, such a prioritization should also group results into topics, or collections of topics made for better search and discovery (see chap. 4). Furthermore it would be helpful, if the correlations would be put in context. That means that the mentions itself should be segmented in different, for the communication operatives' comprehension, dimensions. Likewise this favors a prioritization of the mentions. This prioritization is important by the selection process of scanned social media mentions to collections of mentions to monitor afterwards. Besides, a more precise forecasting can also be supported by the inclusion of context.

Context for social media mentions can, for example, be composed of who said what, when and where. This context understanding is borrowed from journalistic guidelines [Hohenberg, 1978]. Journalism has created a methodical framework guiding the research efforts. The famous wh-questions compose what exhaustive research needs to be able to answer, especially the first four (i.e. what, when, where, and who) being essential and basic. [Dey & Abowd, 2000; Dey et al., 2001] pointed out that activity, time, location, and identity in practice are more important than other types of context to determine why a situation is occurring. Hence, related and interpretative questions of why and in which way, or how, an event has occurred might not always be definable and their answers try to heighten the comprehensibility for the communication operatives. Depending on how one looks at it, context hardly can be determined precise. However, often the combination of presentation of fuzzy context and communication operatives grasp enables to comprehend it better.

Another helpful point for prioritization and ensuing forecasting would be a sentiment analysis of the found mentions. Analyzing the outpouring of millions of prosumers can unveil attitudinal shifts that are not apparent to stakeholder interviews as opinion polls, survey takers, or customer satisfaction surveys (see chap. 4). Tracking public sentiment over time provides invaluable insight and gives the chance to stay right on top of changes in the marketplace and the organization's reputation equity [Sterne, 2011]. So generally speaking, sentiment analysis aims to determine the attitude of a writer with respect to some topic or the overall tonality of a document. A communication operatives' attitude may in this manner be a judgment or evaluation, affective state (i.e. the emotional state of a blogger), or the in-

tended emotional communication (i.e. the emotional effect a blogger wishes to have on the reader). The rise of social media elements has fueled the organizations interest in sentiment analysis. With the proliferation of reviews, ratings, recommendations and other forms of online expression, online opinion has turned into a kind of virtual currency for an organization looking to market their products, identify new opportunities and manage their reputations (see chap. 4). As organizations look to automate the process of filtering out the noise, understanding the conversations, identifying the relevant content and addressing it appropriately, many are now looking to the field of sentiment analysis.

Such a sentiment analysis could support communication operatives in distinguishing social media mentions according to three states for example: Negative, neutral, and positive mentions. One of the tenets of artificial intelligence (and machine learning) is that computers can learn from its deficiencies without having to be told what the deficiency was. It does, however, need to be told that an error occurred (see chap. 3). Over time, it assigns sufficiently mathematical values to a wider variety of errors and makes chance evaluations to determine future decisions. This is just how [Go et al., 2009] teaches computers the word-inherent fuzzy connotations. Though, through such a simple trivalent prioritization first the negative mentions (e.g. for reputation issue and crisis management) can be treated, followed closely by the positive mentions (e.g. to build and stimulate positive communication).

At any rate the communication operatives would consider the Web originators semantic use of vague natural language. Therefore the application should be NLP-able by some means and that way support the communication operative's endeavor. In the next section the integration of the identified online reputation issues is illustrated.

5.3.3 Edit Found Mentions

To support communication operatives in their online reputation analysis process, social media mentions need not only to be found and analyzed according their affiliation, context and sentiment, but also to be stored for a later processing in the ensuing monitoring step. This simplifies the communication operatives' work since they do not need to search for same issues again and again. Furthermore, if several communication operatives work together, they can share these stored reputation issues. Consequently the storage explicitly supports work sharing. A communication operative could, for example, be concerned with scanning for reputation issues and another for the ensuing monitoring. Because the search can be stored, it is possible to pass it from one to another. Moreover, based on these stored issues, they can perform a wider forecasting together. Another aspect of storing information is coming from a controlling perspective. Since for a controlling it is

essential to track made decisions, for example, these stored searches can be used as proof afterwards.

Another important need of communication operatives is to trigger mentions concerning an organization, for example. Accordingly, an application should provide a notification function. This notification function can include a Web search known in advance. For example this could be known weaknesses or mentions about an organization in general (see chap. 4). This known search could be an organization name, brands, service, products and executives personal. Here a trigger function can provide remedy.

A last point is reporting to the management. Mostly a communication operative's task includes a reporting for what reasons the handled reputation issues must be prepared. Hence a function to download entire graphical reports or at least lists (e.g. spreadsheet) with the found mentions is essential as communication assistance for dealing with an organizations' management. With the expansion of information systems, and the desire for increased competitiveness in organizations, there has been an increase to produce unified reports, which join different views of the organization in one place. Consequently this online reputation analysis reporting should be included in an integrated organization report.

5.4 Implications for the Framework

Depending on the previous communication operative's requirements, here the first technical implications for the FORA framework are drawn. These implications are composed of functional requirements to the framework. Hence, predominantly these functional requirements are now viewed from information systems or computer science perspective, but over all they are concerned with the handling of human fuzziness in semantic. Derived from the communication operatives' requirements, in short, the following seven foci are considered important for the framework:

- *User interaction*: A user interface is the space where interaction between communication operatives and computers occurs and interaction accepts and responds to input from humans (e.g. data or commands). Hence the FORA framework should provide an interactive dashboard for vague human-oriented interaction with the communication operatives. Through a segmentation of retrieved Web links, according to context dimensions and an interactive graphical visualization of fuzzy grassroots ontology-inherent knowledge, the user interface facilitates Human-Computer Interaction (HCI). Through knowledge representation this interaction should ease communication operatives reasoning concerned reputation issues.

- *Knowledge representation and reasoning*: The fundamental goal of knowledge representation and reasoning is to represent knowledge in a manner that facilitates drawing conclusions. In other words, it analyzes

how to use symbols to represent a domain of discourse, along with functions that allows formalized reasoning about objects. Consequently the FORA framework must support communication operatives in representing knowledge in a manner that reasoning is possible (e.g. using Topic Maps; see chap. 4). In addition, the framework should back them with an integrated reasoning component that automatically identifies the operatives' needs. This reasoning component should necessarily include context interpretation also.

- *Context interpretation*: Context is defined as any information that characterizes a situation related to the interaction between communication operatives, the FORA framework, and the Social Web, for example. During a search, context can help communication operatives to identify real threats to an organizations reputation. So the FORA framework should support the communications operatives to perceive context. Therefore, some kind of artificial intelligence has to be implemented in the framework to enable an automatic identification of context.

- *Artificial intelligence*: This imitated intelligence is defined as the design of intelligent agents (e.g. using the machine learning technique of fuzzy clustering). An intelligent agent is a system that perceives its environment and takes actions that maximize its chances of success. The FORA framework depends on Web agents that are used to create a copy of visited sites for later processing by a reputation search engine. Another important functional requirement for the FORA framework is the aggregation of vague information.

- *Aggregation of information*: Using Web agents, Web search engines store information on the visited websites. The information of each site is analyzed, and the results of the analysis are stored and indexed for later rapid searching. On the basis of an index, a Web search engine afterwards provides a listing of best-matching websites according to a search query. The FORA framework rests upon the principles of Web search engines and agents. Based on a communication operatives search query, the framework connects the queries with underlying Web content. Thereby the framework stores this in a graph database (i.e. collection of instances and the ontology).

- *Storage*: Knowledge structures (i.e. graphs) grow to enormity at a breathtaking pace. Consequently, to store an entire knowledge base, a flexible knowledge administration system is needed. Such a flexible administration system promises to be a graph database. A graph database is a kind of database that uses graph structures with nodes, edges, and properties to represent and store information. The FORA framework is based on a modern, high-performance, persistent RDF graph database that uses disk-based storage, enabling it to scale to billions of triples while still maintaining good performance. An integration of different da-

tabases techniques (e.g. graph, object, relational, etc.) is strongly desired since every database technique boasts its own pros and cons. This extensibility and scalability are further hot topics of the framework.

- *Extensibility and scalability*: Extensibility on one hand is a design principle where the implementation takes into consideration future growth. It is a systemic measure of the ability to extend a system and the level of effort required to implement the extension. Extensions can be put into effect through the addition of new functionality (e.g. inclusion of new social media elements or integration of additional Semantic Web layers) or through modification of existing functionality (e.g. improvement of interaction with current encompassed social media elements or Semantic Web layers). Scalability on the other hand is the ability to handle growing amounts of work in a graceful manner or its ability to be enlarged to accommodate that growth. A system, whose performance improves after adding hardware, proportionally to the capacity added, is said to be a scalable system. Through the FORA framework's modular structure and the underlying disk-based knowledge administration system the framework enables extensibility and scalability.

These seven foci perceive the FORA frameworks starting situation. They are all important in extracting relevant information concerning an organizations reputation. On the basis of these findings, the following chapter presents the FORA framework in more depth.

5.5 Further Readings

An introduction to the scenario techniques methods offers [Chermack, 2011]. Scenario planning is a tool for surfacing assumptions so that changes can be made in how decision makers see the environment. Learning how to see a situation complete with its uncertainties is an important ability in today's world. This is closely related to case studies. [Flyvbjerg, 2006; Thomas, 2010; van Aken, 2004] help dealing with the numerous challenges of this research method. Their goal is to introduce how to design good case studies and to collect, present, and analyze facts objectively. Lastly, [Hay, 2002]'s compendium describes various requirement analysis techniques.

The introductory Apple Inc. scenario is half borrowed by [Beal & Strauss, 2008] as well as by [Portmann et al., 2012]: From the former the description of the iPhone case is obtained, whereby the fictional scenario of online reputation analysis spring from the latter. The three cases studies have been developed together with postgraduate students. Their Licentiate and Master thesis [Fuchs, 2010; Hächler, 2010; Uhlmann, 2011] are worth reading. The found requirements of communication operatives concerned with online reputation analysis are to a large part described within [Uhlmann, 2011]'s Master thesis.

Part III

Framework and Implementation

FUZZY ONLINE REPUTATION ANALYSIS FRAMEWORK

"Wisdom is knowing what to do next;
virtue is doing it."

—David Starr Jordan

In the Social Semantic Web an organization, a brand, the name of a high-profile executive, or a particular product can be defined as the hodgepodge of all online conversations taking place around it, and this is happening regardless of whether or not an organization participates in the conversationscape's dialogue. Long story short, organizations in the first place are forced to listen to the Social Web so in order to take part in and, in this way, improve their online reputation. To do that intuitively, the FORA framework is conceptualized as a pertinent listening application. So, the term FORA originates from the plural form of forum, the Latin word for marketplaces [Portmann et al., 2012]. Thus, the framework allows organizations' communication operatives a fuzzy exploration of reputation in online marketplaces. Listening and then increasing engagement within social media elements is a hard task. There is a constant flow of information and many organizations do not know how to harness and gain actionable insights from this rich source of customer conversations. The idea beyond the conceptualization of the framework is to listen and in doing so automatically identify key social media elements 24/7 to simplify online reputation analysis and, by that, impart onto communication operatives insightful information on which they can actually act upon. To make this system reality, a design science approach is pursued.

In contrast to natural science, which is a body of knowledge about some class of things in the world that describes and explains how they behave and interact, a design science (aka science of the artificial) is a body of

knowledge about artificial things designed to meet certain desired goals. [Simon, 1996] frames design sciences in terms of an inner and an outer environment, and the interface between the two that meets certain desired goals. The outer environment is comprised of external forces and effects that act upon the artifact. The inner environment involves components that make up the artifact and their relationships to the artifact. The behavior of the artifact is constrained by both the external forces and its components. The bringing-to-be of an artifact, components and their interaction, which interfaces in a desired manner with its outer environment, is the design activity [Vaishnavi & Kuechler, 2009]. Hence, the artifact is structurally coupled to its environment. Anticipated outputs from design research are innovative artifacts such as frameworks (or prototypes).

This chapter completely presents up the FORA framework that can be used to design online reputation analysis applications. At that, the reputation management process in the Social Semantic Web represents the outer environment and the FORA framework architecture and components the inner environment. The interface is a particular component of the framework designed as a GUI (i.e. dashboard application) that allows communication operatives to identify insightful Web information. Inspired by concepts commonly found in conceptual frameworks, some abstractions that help infer higher level information from social media elements and support separation of concerns have been defined. In the following sections, these derived abstractions for online reputation analysis are defined and the how to derive these abstractions from the application specifications are described. Section 6.1 first presents in a nutshell the concepts behind the FORA framework. In clarifying these ideas, section 6.2 reveals the frameworks architecture and components' interaction. Subsequently, in section 6.3 the framework's key components are compared. Afterwards section 6.4 showcases the implications for an implementation of these key components. Finally, section 6.5 references further readings.

6.1 Outline of the Framework

In the previous chapter, several aspects of online reputation management were presented. The importance of a robust online reputation management involves the communication operatives' assignment of online reputation analysis. The communication operatives' needs for performing online reputation analysis have been detailed through a scenario and three case studies (i.e. react to mentions, put mentions in context, and edit found mentions; see chap. 5). On this basis, the most important functional requirements for a smoothed and at the same time enhanced online reputation analysis are inferred. Yet, all these requirements have to be integrated into a conceptual framework that defines why someone or something is doing something in a particular way. Putting together the FORA framework from the investigated requirements is a unique process. In conformity with [Maxwell, 2005], the main goal involved a need for integration of the single components with one another, and with the research questions. The FORA framework thereby eventuates in a modular overarching architectural model.

The design of this architecture raises challenges. Online reputation analysis in the Social Semantic Web is difficult for at least the integration of the functional interaction of the seven research foci (see chap. 5). Nothing but an interactive dashboard applet for a sound communication operative's user interaction allows umpteen implementation options. According to [Hearst, 2011] information seeking can be seen as being part of a larger process of sense making that may involve in a larger part also fuzziness. To achieve a good implementation, it is (sometimes) useful to go adrift of similar realizations to achieve a good social acceptance (see chap. 2). The dashboard incorporated knowledge representation can correspondingly be implemented in many different ways but here, by contrast, it is at least possible to orient on (global) standards or at least on common views [Pepper, 2010; van Harmelen et al., 2007].

Context is crucial to properly judge reputation-relevant situations. Usability studies indicate that displaying the query terms in the context in which they appear in the document improves the user's power to gauge the relevance of the result [Hearst, 2011]. Typically it consists of the location, identity, and state of people, groups, and computational and physical objects. Since these may include fuzziness, context interpretation can have many pitfalls because context-perception can vary [Dey & Abowd, 2000; Dey et al., 2001]. With artificial intelligence (and machine learning) this context perception can be simplified [Bishop, 2007; Lämmel & Cleve, 2008]. Fully automated Web agents that first aggregate and evaluate Web information and a subsequent intelligent representation of the found information can endorse communications operatives' senses for context. These senses are, in contrast to computers, inherent in humans—what for a computer may be difficult to analyze, for a human may not be.

To store the fast-growing information found by the Web agents corresponding storage space should be made available. However, this is not an easy undertaking and scientists are still working to develop optimal solutions to adapt to fast-growing data. Nevertheless, if the principles of Atomicity, Consistency, Isolation, and Durability (ACID) are ignored, then most graph databases seem to scale well as data grow and in many cases they are flexible enough to shelter semi-structured and sparse Web data [Tiwari, 2011]. This way, the system can readily scale up, for example by adding more servers, and failure of a server can be tolerated. Truly current graph databases fall short if operations require many complicated structures. Because of this, a sound mix of different extensible and scalable database techniques (see chap. 5) would be an ideal solution. Likewise the respective storage management tool should be extensible and scalable in an easy way [Tiwari, 2011; Bondi, 2000]. In our case, for the FORA framework, the underling management tool should be replaceable modularly to react on scientifically amended insights.

When looking at a new technological advance and when trying to make it easily accessible, a common stepping-stone is to apply tried-and-true software engineering principles to the particular case of the new technology at hand. The acceptance of real-world vagueness and the incorporation of fuzziness facilitate a more natural interaction with social media elements. Awareness of such online reputation issues opens new possibilities that can drive the development of new adaptions that can even be beneficial to legacy systems. Online reputation analysis applications need to take advantage of all social media elements available. Thus, the major preoccupation is to achieve a separation between the system per se and information acquisition and processing. On that account, for the FORA framework architecture, some abstractions that facilitate the inference of higher-level information from social media elements and encourage the separation of concerns have been defined. Together with the identified requirements and the derived functional interaction (see chap. 5) that resemble the extensions of traditional GUI capabilities, the following section introduces the FORA framework's architecture and its component interaction. In the course of this, it is explained how to derive abstractions from the specifications.

6.2 Architecture and Component Interaction

The purpose of software architecture is cutting complexity through abstraction and separating of concerns. As a maturing discipline designing software architecture is a mix of art and science. The art aspect arises out of the various requirements to a corresponding application.

In chapter 5 the requirements for fuzzy online reputation analysis were inferred. Subsequently, the FORA framework architecture provides a mapping of communication operatives' requirements with validation tests. Moreover, the FORA framework architecture comes over as a partitioning scheme,

which partitions the application's current as well as foreseeable requirements into well-defined components. It is a partitioning scheme, which is exclusive, inclusive, and exhaustive. A purpose of this partitioning is to structure the subsystems elements so that there is a minimum of communications needed among them.

The FORA framework architecture is the conceptual model that defines (among others) the structure and behavior of the online reputation analysis application. It permits searching the Social (Semantic) Web to find sources on organizations' online reputation. Using this framework, it is possible to scan the Web according to a query, in order to determine topic classes with related tags and, thus, to identify hidden information.

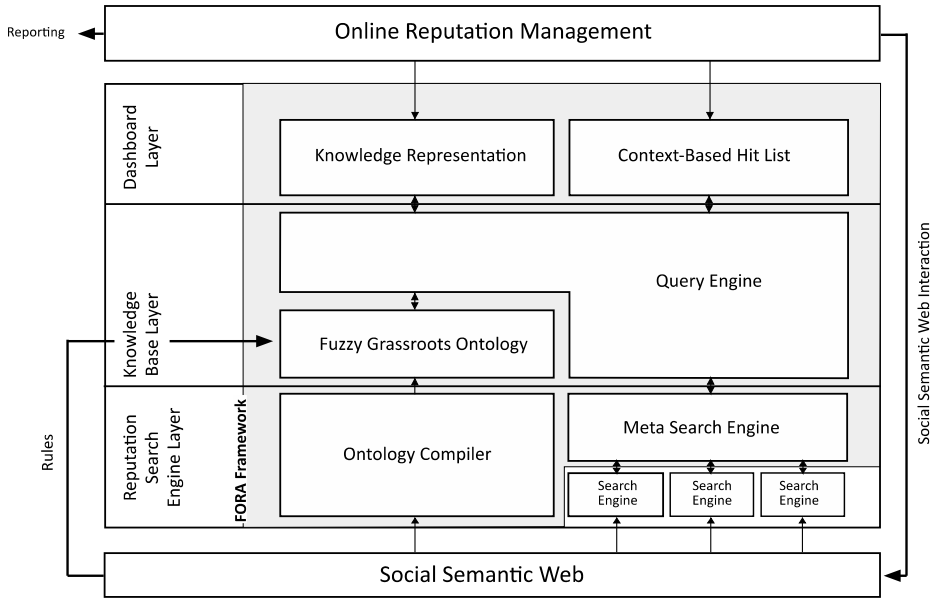

Figure 6.1: FORA Framework Architecture.

Figure 6.1 illustrates the FORA framework's modular structured rough architecture. This architecture consists of three key layers: the reputation search engine layer, the knowledge base layer and the dashboard layer. All these layers contain applicable components. For instance, the dashboard layer contains the knowledge representation and the context-based hit list components. The knowledge base layer contains the query engine and the fuzzy grassroots ontology. The ontology compiler and the metasearch engine components belong to the reputation search engine layer. Besides the notion of the components, the point is also to demonstrate their interaction that should fit together like gears. In order to do that, the flow of information is followed (see fig. 6.1). For this purpose the start and end point is the Social Semantic Web used for online reputation management. The three layers discussed are:

- *Reputation search engine layer*: The reputation search engine layer is designed to search for information on the Social Semantic Web. To this end the reputation search engine avails Web agents. It consists of two components: First, an ontology compiler that algorithmically collects metadata and converts it into the already familiar fuzzy grassroots ontology (see chap. 3). Second, a metasearch engine that sends communication operatives' search requests to several subjacent search engines (i.e. not part of the FORA framework) and aggregates their search results (i.e. hits) into a list. However, the metasearch engine, which operates on the premise that more comprehensive hits can be obtained by combining the results from several search engines, splits the found hits into context-based hit lists. In other words, the metasearch engine distributes the found hits into hit lists according to context dimensions (e.g. identity, location, status (or activity), and time). In the course of this, the metasearch engine mediates between the various search engines (i.e. outside the framework) and the query engine and creates also a ranking in the hit lists.

- *Knowledge base layer*: If instances are annotated and modeled as an ontology, then the talk is of a knowledge base (see chap. 2). A knowledge base is a collection of instances of the concepts defined by the fuzzy grassroots ontology that specifies the structure of the knowledge stored. The knowledge base contains the familiar fuzzy grassroots ontology (see chap. 3), and a query engine. Upon a closer look, figure 6.1 reveals that the fuzzy grassroots ontology (i.e. the technically beating heart of the knowledge base) is in this layer surrounded entirely by the query engine. This query engine is the linchpin between the reputation search engine layer and the dashboard layer.

- *Dashboard layer:* An interactive information system GUI that is designed to be easy to read. The dashboard is the frameworks layer, used for hosting two mini-applications (i.e. components): One is the knowledge representation, the other the context-based hit list. Technically these widgets are bound together by the query engine, which for one thing translates entered communication operatives need into a proper SPARQL query. Thereby the query engine rummages around the fuzzy grassroots ontology and also conveys the fuzzy grassroots ontology-enhanced search query (see chap. 4) to the metasearch engine. In other words, the query engine translates the search term to SPARQL in order to query the fuzzy grassroots ontology. The extracted fuzzy grassroots ontology knowledge is visualized on the dashboard in the interactive knowledge representation widget. Simultaneously the query engine delivers the fuzzy grassroots ontology enhanced search query to the metasearch engine, which in turn, charges various search engines with the retrieval. Thereafter the metasearch engine splits the search engines' hits into fuzzy context dimensions and redelivers them to the query engine again.

Through the context-based hit list widget, the query engine displays the found hits to the organization's communication operatives.

In the sense of a closed online reputation management, the located reputation issues must be stored and notifications must be allowed to be triggered. By this means it is possible to feed the organization's reporting system. However, to close the online reputation management loop, the ascertained reputation issues must be addressed and subsequently controlled. In the following sections all the previous overview, briefly described layers, with their associated components, are explained in more detail. In so doing, the realization (i.e. the "how?") of the particular components of the FORA framework is emphasized. Section 6.2.1 starts with the reputation search engine layer, then section 6.2.2 introduces the knowledge base layer, and last but not least, section 6.2.3 highlights the dashboard layer in more depth.

6.2.1 Reputation Search Engine Layer

The reputation search engine layer in general rests upon the principles of Web search engines. Using Web agents this search engine stores information (e.g. URL and metadata) on the visited websites. The information of each site is analyzed, and the results of the analysis are stored and indexed for rapid searching later. Based on an index, the Web search engine later provides a listing of best-matching websites according to a search query [Manning et al., 2008; Baeza-Yates et al., 2011]. The reputation search engine is a special case of Web search engine that sends a given query to several search engines, collect their replies and pool them in a single ranked list. As illustrated in figure 6.2, it can be seen as a type of federated search where the federated sources are independent search engines. Note that in this figure the reputation search engine contains both the reputation search engine layer and also the knowledge base layer.

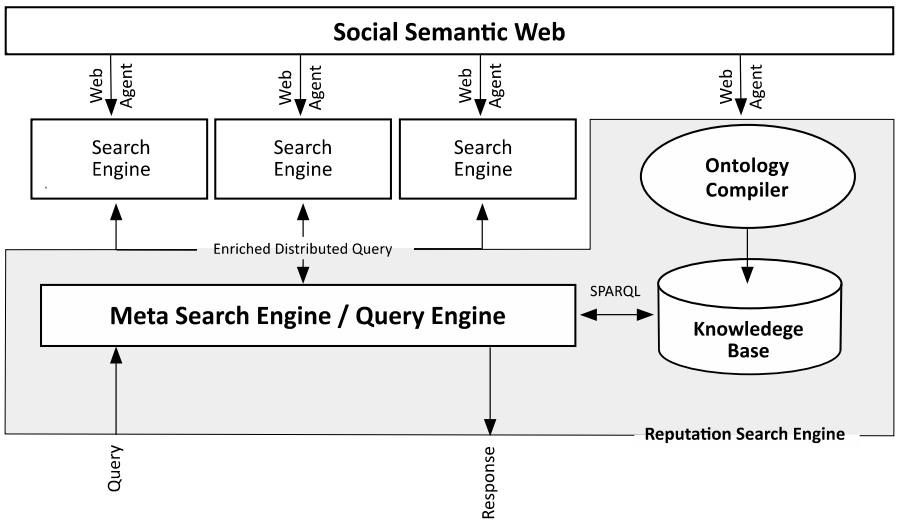

Figure 6.2: Reputation Search Engine.

The reputation search engine's advantage is that it allows for one thing sorting hits by different context dimensions (see chap. 5), which can be more informative than the output of a single search engine [Baeza-Yates et al., 2011]. Based on the characteristics of social media elements, they can be allocated to [Dey & Abowd, 2000; Dey et al., 2001]'s four minimal context dimensions, for example. To this effect it is feasible to use weights (e.g. fuzzy membership degrees) for allocating an appropriate element to a dimension. The higher a weight the greater a specific element belongs to the dimension and, in the course of this, the further forward in the hit list. The found individual entries of the respective social media element follow the same priority setting. Any duplicates are sorted out during this process by comparing their URLs. However, the allocation needs to be set up manually for the first time but based on machine learning, it could be enhanced to an automatically allocation of new social media elements to the appropriate dimensions [Bishop, 2007]. Additionally, the reputation search engine search corpus is dynamically built and thereby the relevant Web contents can be spotted at run time.

The ontology compiler, however, is built on the concept of Web agents that constantly crawls the Social Web, looking for tags. In fact, the ontology compiler relies not only on one but also on several Web agents. These agents identify all tagged sources in folksonomies and subjoin them into a list. During this process the tags are normalized because of a well-known problem in folksonomies where people choose their own tags to annotate sources: typing errors that can occur since there is no editorial supervision. This leads to overlapping but barely relating terms in the underlying ontology. Certainly it can be assumed that a search system finds relevant information despite misspelling in tags, because the queries contain the same mistakes. But the necessity of fault-tolerant treatment of queries becomes clear [Lewandowski, 2005]. Tag normalization is a process by which tags are transformed to make them consistent in a way in which they were not before. This normalization is performed before the tags are further managed.

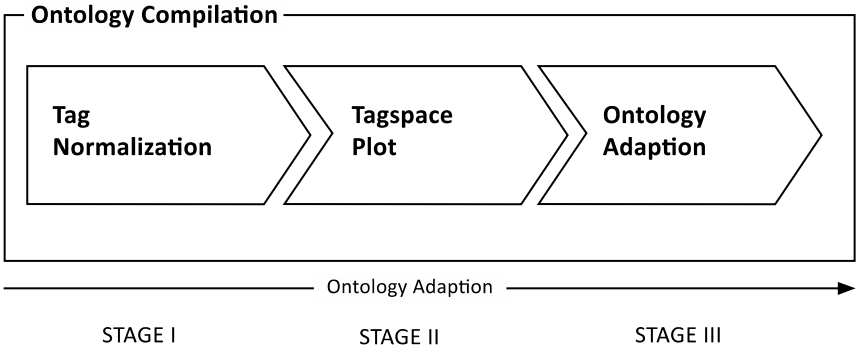

Figure 6.3: Ontology-Compilation Process.

Figure 6.3 illustrates the fuzzy grassroots ontology compilation process. The first step is concerned with the aforementioned tag normalization. To normalize the Web agents found tags, first the ontology compiler converts all of them to upper or lowercase. Comparing the word-inherent text strings identify homographs at that. A fully automated phonetic comparison is performed to detect homophones—the second part of the homonyms (see chap. 3). For this, in a comparison similar sounding tags are encoded to the same code, so that they can be matched despite minor differences in spelling. Examples are bow and bough or Mrs. Powell Job's (i.e. the widow of Apple founder and former CEO Steve Jobs) first name Laurene, that has many different alternatives as Lauraine or Laureen which are pronounced the same. Searching for one spelling would not show results for the two others. Since by using a phonetic algorithm similar sounding words share the same code, here all three variations produce an identical one. This code can then be compared with a directory of words with the same or similar code. Words with same or similar codes then become possible alternative spellings [Baeza-Yates et al., 2011]. Lastly, using a lexical database that groups English words into synsets, helps identifying synonyms in the collected tag set. A synsets thereby is defined as a set of synonyms that are interchangeable in some context without changing the truth-value of the proposition in which they are embedded [Saeed, 2008].

Then, the second step to compile the fuzzy grassroots ontology is concerned with the creation and plotting of a tagspace. For that to happen, the previously collected and normalized tags are linked to each other using a metric function that compares the distance d of the accumulated tags (see chap. 3). Tag similarity is measured as a kind of semantic correlation between tags, considered by means of relative co-occurrence among tags [Hassan-Montero & Herrero-Solana, 2006; Kaser & Lemire, 2007; Budura et al., 2009]. That is, relative co-occurrence is identical to the partition between the amount of resources in which tags co-occur, and the amount of resources in which any one of the two tags appear. As presented in chapter 3, using variants can retain issues. Yet, this collection method causes these tags to become united and offers a semantically consistent picture where nearly all tags are related to each other. In the course of this, finally the synonyms (i.e. often tagged together) can be matched with the lexical database, thereby allowing a more accurate or fine-grained determination of synonymity. The resulting semantically consistent picture is referred to as tagspace (i.e. a distance matrix). After this step, all of the tags are linked to each other and—based on [Bourke, 1997]'s intersection of two circles—plotted onto a tagspace, which is the input for the ontology adaption. This is the compliance of [Hearst, 2011]'s concept of placing tags on a two-dimensional canvas.

The ontology adaption is the last step of the fuzzy grassroots ontology-compilation process that separates the plotted tagspace into different clusters with the help of a fuzzy clustering algorithm. By minimizing an objective

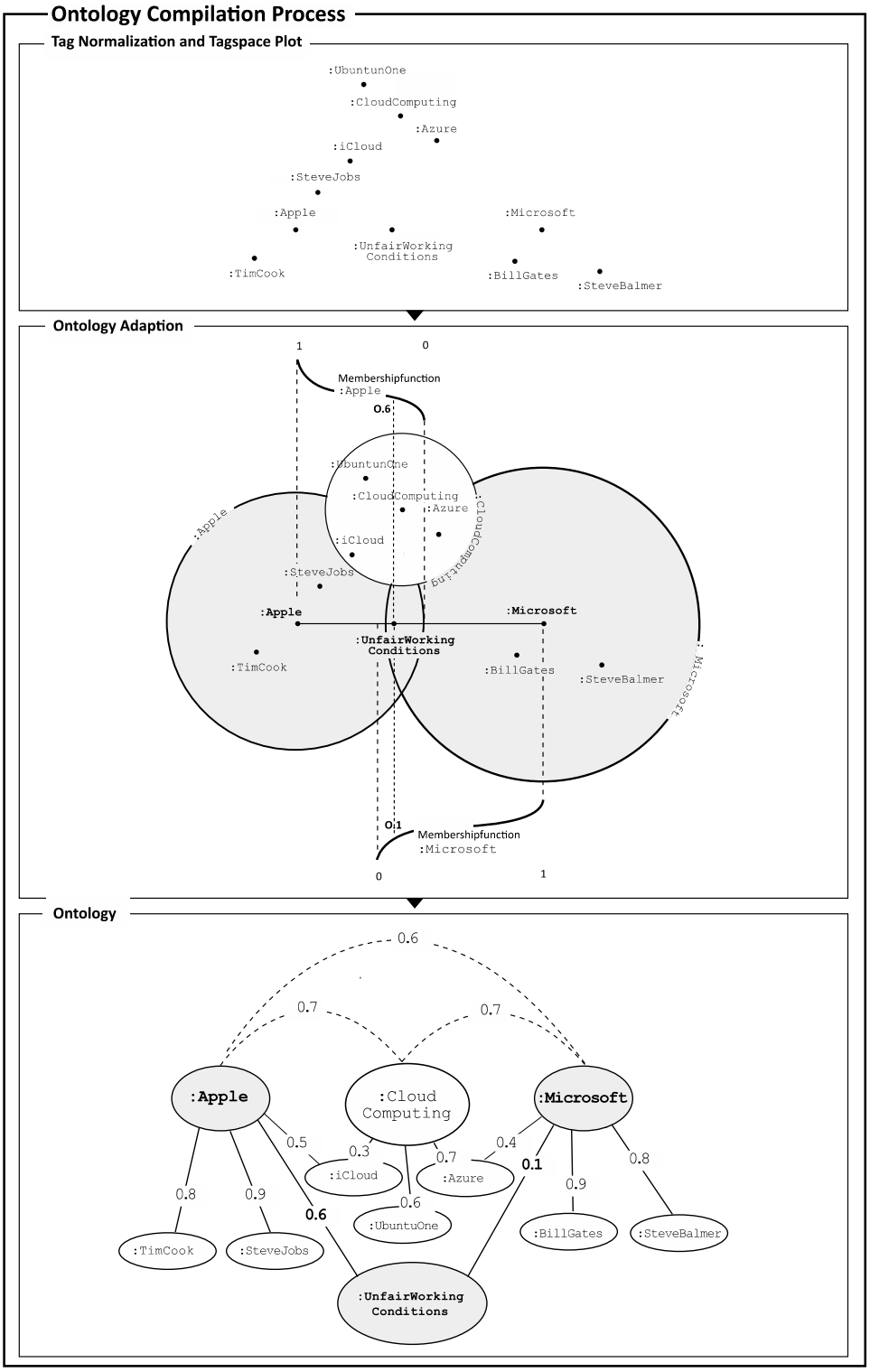

Figure 6.4: Ontology Compilation Example.

function J_f such an algorithm partitions a finite collection of n elements in the dataset X into a collection of c fuzzy clusters with respect to some given criterion. Given a finite set of data, the algorithm returns a list of c cluster centers $C = \{C_1, \ldots, C_c\}$ and a partition matrix $U = u_{ij} = \mu_{\Gamma_i}(x_j) \in [0,1], i = 1, \ldots, n, j = 1, \ldots, c$, where each element u_{ij} tells the degree to which element x_j belongs to cluster Γ_i [Bezdek, 1981](see chap. 3). The tag nearest to the center names the class, and the other tags—including the eponym itself—are stored in this class by name and the membership degree for belonging to the class. To this end, three common different fuzzy clustering algorithms are contrasted in section 6.3.1.

Figure 6.4 sketches out the presented ontology-compilation process by way of a well-known example. The Web agents in folksonomies automatically found tags around Apple, cloud computing, and Microsoft are plotted on a tagspace. Using fuzzy clustering these tags of the tagspace are adapted to fuzzy clusters (i.e. `:Apple`, `:CloudComputing`, and `:Microsoft`), with the tag nearest to the cluster center naming the class. The resulting fuzzy grassroots ontology consists of the classes and containing elements, their relationships, and properties such as the degree of relationship (i.e. fuzzy membership degree u_{ij}). The membership functions in figure 6.4 are arbitrarily chosen to illustrate the idea of the ontology adaption. Yet, [Kaufmann & Meier, 2009] for example, present a possible supervised classification for membership function induction based on normalized likelihood ratios. This discrimination, in future could be used to enhance the ontology compilation process and thereby make it even more adaptive and naturally.

The FORA frameworks metasearch engine is a search engine that sends the fuzzy grassroots ontology-enriched query to several subjacent search engines and aggregates the results into lists of context. However, the enrichment of the user query is based on the knowledge base administrated fuzzy grassroots ontology. The fuzzy grassroots ontology is queried using SPARQL. Next, the metasearch engine takes the enriched query, passes it to several other heterogeneous search engines and then compiles their results in a homogeneous manner. To this end, the metasearch engine decomposes the enriched query into subqueries for submission to the constituent search engines. Because various search engines employ different query languages, the metasearch engine applies wrappers to the subqueries to translate them into the appropriate query languages of the underlying search engine [Baeza-Yates et al., 2011]. Last, the metasearch engine collects their replies and classes them with context dimensions.

6.2.2 Knowledge Base Layer

The knowledge base layer is a collection of instances (i.e. ABox) of the concepts defined in the fuzzy grassroots ontology, and the fuzzy grassroots ontology (i.e. TBox) specifies the structure of the knowledge stored in the knowledge base. Together, ABox and TBox make up the entire fuzzy grass-

roots ontology-based knowledge base [Kaufmann, 2007] that is the beating heart of the FORA framework. Since the framework is based on the reputation search engine, the knowledge base represents the fuzzy grassroots ontology plus the ad hoc found Web contents.

An ontology is the core of the FORA framework. Using the presented Semantic Web technologies (see chap. 2), it is stored in a knowledge administration system. The last subcomponent of the ontology compiler separates the plotted tagspace into classes with the help of a fuzzy clustering algorithm spawning the fuzzy grassroots ontology. Based on a high-performance and persistent graph database, the underlying knowledge administration system provides a solid storage layer (see chap. 2) with reasoning capabilities, which are yet not used. In contrast to other database system (e.g. relational or object), a graph database bears each stored item in mind to have any number of relationships. These relationships can be viewed as links, which together constitute a graph [Tiwari, 2011]. The core storage uses disk-based storage, enabling to scale to billions of triples: yet, the focus of this PhD thesis is not on databases and their hardware. In fact, the entire knowledge administration system is designed to manipulate triples that can be queried with SPARQL. To this end, three different reasonable knowledge administration systems are compared in section 6.3.2.

The query engine is a crucial point of the FORA framework and acts as hub between the fuzzy grassroots ontology, the metasearch engine and the dashboard. It translates an entered user query to a SPARQL query and sifts through the fuzzy grassroots ontology. Afterwards, it translates the enhanced query to the metasearch engine that decomposes the enriched query into subqueries for submission to the underlying search engines. Last it provides the dashboard with the classified replies collected from the metasearch engine.

6.2.3 DASHBOARD LAYER

The query engine conveys the context classified to the dashboard layer, which is designed as a small application (i.e. an applet) that performs online reputation analysis within the scope of the FORA framework application. In this framework, the dashboard applet (i.e. a small Web-based JavaScript application) does not run independently; in fact it must run in a repository provided by the application. In response to the communication operatives' interaction, the applet changes the provided (graphically and textually) content. Furthermore the dashboard applet provides various functions such as storing, triggering and reporting search results. Nevertheless, its key elements are two widgets: The knowledge representation and the context-based hit list widgets.

As previously introduced (see chap. 4), knowledge representation is the field which is concerned with the evaluation of data to identify relevant concepts, relations, and assumptions. Its aim is to present data in a manner that will

enable reasoning. On one hand, the fuzzy grassroots ontology provides computers with a general knowledge of vague human concepts and, on the other hand, the fuzzy grassroots ontology-based interactive visualization of this knowledge through a visualization technique helps people to identify patterns. With this in mind, in section 6.3.3 three different knowledge representation systems are contrasted. Thereby Topic Maps are selected as good solutions to visualize knowledge. Figure 6.5 sketches out the ontology and its visualization as Topic Maps again of the famous Apple-Microsoft example. Importantly, for a straightforward search, this interactive visualization can be used as a starting point.

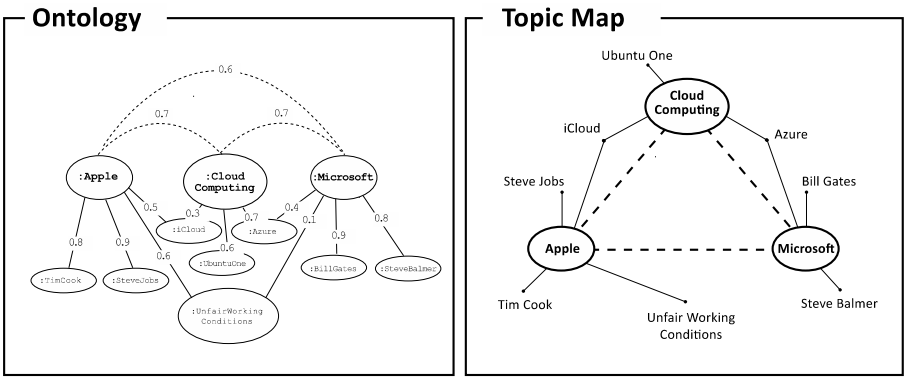

Figure 6.5: Ontology Visualization Example.

Using fuzzy clustering, the Apple-Microsoft tags of the tagspace were adapted to fuzzy clusters. Following an algorithmic process, the class is named automatically and each tag is likewise assigned to one or more classes (see fig. 6.4). The resulting fuzzy grassroots ontology consists of the classes and including elements, their relationships, and properties. From figure 6.5 it can be learned that for example the Class `:Apple` is related to the class `:CloudComputing` with a degree of relationship (i.e. the fuzzy membership degree) of 0.7, and with `:Microsoft` with a degree of 0.6. The Class `:CloudComputing` consists of the elements `:Azure`, `:iCloud`, and `:UbuntuOne`. The degrees of relationship for this example are arbitrarily selected. On the Topic Map, the length of the edges (i.e. based on the distance d) now visualizes these membership degrees: Apple and Microsoft have exactly the same distance from cloud computing, whereas the distance from Apple to Microsoft is larger (i.e. less close related). The elements Azure[31], iCloud, and Ubuntu One[32] have varying distances to cloud computing. Azure has a membership degree of 0.7, Ubuntu One 0.6, and iCloud 0.3.

As already presented, [Dey & Abowd, 2000; Dey et al., 2001] introduced four essential characteristics, of context information—identity, location, status (or activity) and time. Identity refers to the ability to assign a unique identifier to an entity. Location is all information that can be used to deduce spa-

tial relationships between entities. The status identifies intrinsic characteristics of the entity that can be perceived. Finally, time is context information as it helps characterize a situation. It enables to leverage off the richness and value of historical information. It is most often used in conjunction with other pieces of context, either as a timestamp or as a timespan, indicating an instant or period during which some other contextual information is known or relevant. This indicates fuzziness.

Bearing this in mind, the proposed context information can be analog displayed vague because they stem from ever-changing Social Semantic Web. There the underlying data change constantly, which is why no simple hard fragmentation to context dimensions, can be made. For example, the dimension identity can consist of fuzzy clusters comprising stakeholder groups (e.g. affiliates, communities, consumers, etc.; see chap. 4). The respective data, for example, can be extracted from the network-like social network structures. At that, the relationship among the stakeholders theoretically can be analyzed. The dimension location can consist of fuzzy clusters as geographical location (e.g. America, Asia, Europe, etc.); the dimension status of tagged tags (e.g. Apple, cloud computing, Microsoft, etc.); and the dimension time of timestamps (e.g. today, yesterday, last seven days, etc.). The respective data may all be extracted from the different social media elements (see chap. 5).

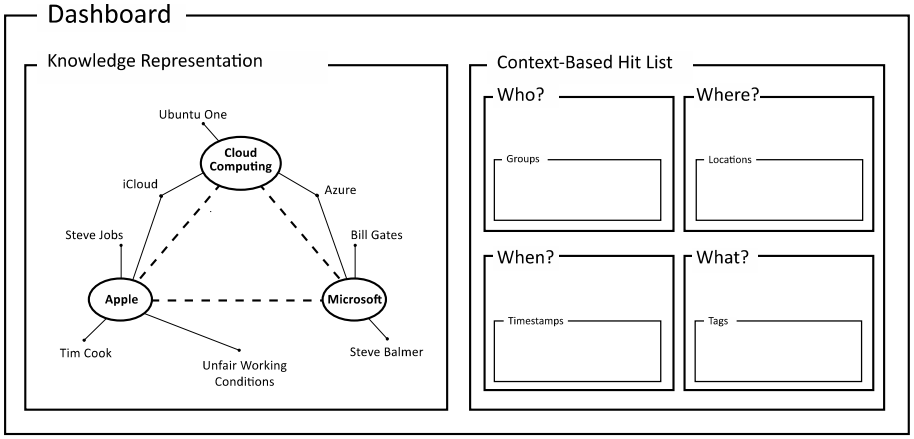

Figure 6.6: Online Reputation Analysis Application Dashboard.

In figure 6.6 the dashboard application is outlined: On the left, the knowledge representation widget is implemented using Topic Maps to represent fuzzy grassroots ontology-inherent knowledge, supplemented on the right by the context-based hit list widget. These basic context categories provide information that can be used to infer additional pieces of context and lead to a more extensive assessment of a situation.

So, what really makes the dashboard unique is what happens after the communication operatives' enter his search term. Instead of delivering millions of search results in one long list, the context-based hit list widget arranges similar results together into context-dimensions and the knowledge representation widget appendant represents interactively the fuzzy grassroots ontology-inherent knowledge. That way it is possible to search results by topic and so a communication operative can zero in on exactly what he is looking for or discover unexpected relationships between items. Rather than scrolling through page after page, the dashboard's widgets help find results that may have been missed or that were buried deep in a ranked list. Thus using the FORA framework, it is possible to analyze the Social Web for a name, product, brand, or combination thereof and determine query-related topic classes with related terms and thus maybe hidden reputation issues.

6.3 Comparisons of Key Components

There are different ways to implement (at least parts of) the FORA framework. This goes hand in hand with the theory of conceptual frameworks that is in particular a built rough-and-ready theory. It covers borrowed components and the overall consistency is artificial and not something that exists out of the box (see sect. 6.1). Thus, implementing the FORA framework allows various possibilities. In this PhD project the stated objectives from chapter 1 (i.e. a comparison of fuzzy clustering algorithms, knowledge administration and knowledge representation systems) make out the chosen ones. The abundance of different algorithms or systems to choose from, frequently causes decision issues. In the following sections for a given situation different solutions were compared in the following sections. Each of these comparisons was collaborative work with respective undergraduate students at the University of Fribourg (Switzerland).

Supported by [Kolly, 2011]'s Bachelor thesis, section 6.3.1 starts with a comparison of three different applicable fuzzy clustering algorithms. To this end, an algorithmic point of view was chosen. For the following two sections in contrast, a system and theory-oriented point of view was elected. Based on [Osswald, 2011]'s Bachelor thesis, section 6.3.2 compares three different knowledge administration systems. Therefore it was relied on W3C recommendations. Last but not least, section 6.3.3 presents an adapted version of [Burkhard, 2011]'s Bachelor thesis presented comparison of three different knowledge representation systems.

6.3.1 Fuzzy Clustering Algorithms

To build a fuzzy grassroots ontology from folksonomy tags, first a normalized tagspace is needed. Likewise in this tagspace all the data are in a form that from every tag its neighbors and the distance to these neighbors are known. This calculation is based on the intersection of two circles on a plane [Bourke, 1997]. Now, the tagspace is ready for a fuzzy clustering. Among the advantages of fuzzy clustering algorithms ranks an improved accuracy of

clustering under vagueness, a relatively straightforward practicability, robustness as well as low solution cost. Yet, in order to elect the most suitable fuzzy clustering algorithm for the FORA framework, with assistance of [Kolly, 2011], a comparison of different algorithms is drawn. Among others, [de Oliveira & Pedrycz, 2007] and [Miyamoto et al., 2008] list various fuzzy c-means-based clustering algorithms:

- *Fuzzy C-Means (FCM):* One of the most widely used fuzzy clustering algorithm is [Bezdek, 1981]'s FCM algorithm. This is one of the oldest fuzzy clustering algorithms, which is why it is rather simple but no less effective.

- *Gustafson-Kessel (GK):* Another one is the GK algorithm that is an extension of the conventional FCM, which is able to detect clusters of different geometrical shapes.

- *Fuzzy clustering by Local Approximation of MEmbership (FLAME*): The two previous fuzzy clustering algorithms derived from the aforementioned books, are now also compared with a similar but newer one, presented by [Fu & Medico, 2007]. Their FLAME algorithm is able to capture non-linear relationships, non-globular clusters and, more importantly, to automatically define the cluster numbers as well as outliers.

To evaluate the most suitable fuzzy clustering algorithm, for an automatic and robust ontology-creation suitable factors are needed. In order to find such factors, all three algorithms are examined for potential comparable properties. Since the algorithms are mathematically comparable, the factors must be measurable and verifiable. This is guaranteed twofold: First through mathematical comparisons, and second by matrix laboratory (Matlab) tests. Referred to as numerical computing environment and fourth-generation programming language, using Matlab the mathematical comparisons were (visually) supported. Thus the found factors for the comparison are complexity, permanence, and adaptability [Kolly, 2011]. Table 6.1 illustrates the selected factors (with dedicated weight and value range).

Table 6.1: Critical Factors to Fuzzy Clustering Algorithms.

Factor	Weight	Value Range
• Complexity	50%	{0; 1}
• Permanance	30%	{0; 0.5; 1}
• Adaptability	20%	{0; 0.5; 1}

The complexity of a fuzzy clustering algorithm may be polynomial at the maximum, because if an algorithm is not processable within polynomial time it is inacceptable and must be rejected as unprocessable and insolvable within practical time [Goldreich, 2008]. Moreover, an algorithm should have certain permanence. Changes should be minimized that not all small adjustments creates a completely new ontology. In addition, also coincidence

should play a small role as possible. Although fuzzy clustering is an unsupervised machine learning technique, if worst comes to worst, it should be possible to adjust an unsatisfactory result by externally modifying a parameter. This denotes the adaptability of the algorithm.

For all the three compared algorithms, the highest achievable complexity is quadratic. Consequently the complexity criterion complies by all algorithms. On that account all three algorithms are valued with 1.

Concerning the permanence the fuzzy clustering algorithms differ. The fact that the number of clusters is attributed at the beginning of FCM and GK algorithms, any variation leads to the creation of another cluster distribution. Since also all memberships are set randomly at the start, the cluster can, besides, look relatively differently at any initialization. In the GK algorithm, however, the clusters are changed less since geometric shapes are recognized. Nevertheless, the FCM and the GK algorithm comply with the permanence criterion only poorly. Not so with the FLAME algorithm that automatically detects the number of clusters based on maximum density. In addition, there are no coincidences and each execution results in exact the same result. Hence, the permanence criterion is met. Therefore the scoring sequence looks like this: The FLAME algorithm is rated with 1, the GK algorithm with 0.5 and the FCM with 0.

A determination of the number of classes in advance can be an advantage as well as a disadvantage. With this parameter a desired output can be selected at the very beginning, what can be regarded as advantage. In an unsupervised machine learning task, such as fuzzy clustering, this parameter can be interpreted as disadvantage too. However, the FCM as well as GK algorithm comply with the adaptability criterion. On these grounds they are both rated with 1. In the FLAME algorithm a change in the class centers is induced only by a major change from the number of closest neighbor around the cluster centers. Yet, an undesired result could possibly be altered. Hence, the adaptability criterion is met as well but the FLAME stays closer to a real unsupervised machine learning algorithm. However, both FCM and GK comply better with this criterion, so the FLAME algorithm is in comparison only rated with 0.5 (see tab. 6.2).

Table 6.2: Comparison of Fuzzy Clustering Algorithms.

Factor	Weight	FCM	GK	FLAME
• Complexity	50%	1	1	1
• Permanance	30%	0	0.5	1
• Adaptability	20%	1	1	0.5
Total	**100%**	**0.70**	**0.85**	**0.9**

The FCM algorithm is based on simple principles and as a result not very progressive. In many cases, this straightforward algorithm is adequate, for the creation of the fuzzy grassroots ontology it is not. The GK algorithm is an

interesting extension of the FCM algorithm. Because of the detection of clusters of different geometrical shapes it is frequently preferred. However, because of the artificial determination of the number of classes in advance, for the creation of the fuzzy grassroots ontology it is not considered. The FLAME algorithm, however, allows capturing non-linear relationships and non-globular clusters and, more importantly, an automated determination of cluster numbers. In addition, it also recognizes outliers. The produced number of clusters is more naturaly since they bear by their density on information entropy; where there is a higher information density, a cluster with associated center can be identified by default. On these grounds for the implementation of the YouReputation prototype the FLAME algorithm is selected.

6.3.2 KNOWLEDGE ADMINISTRATION SYSTEMS

To mount ontologies, special programs are required that facilitate their creation. For this purpose the following section deals with knowledge administration systems. In the recent past, a variety of different systems were developed (mainly by universities). Thereby all producers voiced different needs for the respective systems. Hence, the focus of these systems is relatively different. Many available applications are academic prototypes, meaning that most of the implementation in the query language aims to support but not to provide the necessary programming and administrative abilities to make them operational within a real working environment. Besides the emerging software that supports ontology, an increasing number of ontology applications boost the advancement of storage and query support.

Within this PhD project, no new knowledge administration tool was implemented. Instead it was relied on W3C recommendations. Thus [Portmann et al., 2012] perform a first initial distinction (i.e. list) of prominent knowledge administration systems; this list serves within this thesis as a basis. For the present value analysis of a suitable system for the FORA framework of the knowledge administration systems three different ones of various backgrounds are compared to each other. The comparison includes a system that was created by a professional vendor, a pure open-source system, and a university-created and for general use released system. This broad selection provides an overview of the quality of the different development philosophies (i.e. professional vs. open-source) [Osswald, 2011]:

- *AllegroGraph*: A closed source graph database from Franz Inc., a professional vendor of Semantic Web technologies. Its underlying RDFStore is a high-performance, persistent graph database. It uses disk-based storage, enabling it to scale to billions of triples while maintaining superior performance. Additionally, it supports RDFS, OWL, SPARQL, and Prolog reasoning from numerous client applications.

- *Jena*: A Java framework for building Semantic Web applications originally from Hewlett-Packard research laboratory but later released for the whole open-source community. It provides a programmatic environment for RDFS, OWL, and SPARQL. The Jena framework includes in-memory and persistent storage and includes the Pellet reasoner (i.e. an open-source Java OWL DL reasoner).

- *KArlsruhe ONtology (KAON*): An open-source, Java-based ontology management infrastructure targeted for business applications developed at the University of Karlsruhe (Germany)[33]. It includes a comprehensive tool suite allowing easy ontology creation and management. KAON readily supports RDFS, OWL, and SPARQL. Persistence mechanisms of KAON are based on relational databases.

To evaluate the most suitable knowledge administration system, different factors have to be analyzed. Adapted from [Osswald, 2011], the tables 6.3 and 6.4 illustrate the selected factors; thereby the former shows the critical factors to the FORA framework, the latter the ones that are nice to have. However, implementing these three systems and checking them thereby against each other have analyzed these factors. They were assigned according to their importance for the FORA framework. Subsequently, table 6.3 reveals the critical factors to knowledge administration systems.

Table 6.3: Critical Factors to Knowledge Administration Systems.

Factor	Weight	Value Range
• Number of supported Languages	20%	{0; 0.2; 0.4; 0.6; 0.8; 1}
• Speed of Retrieval	25%	{0; 0.25; 0.5; 0.75; 1}
• Type of Storage	20%	{0; 1}
• Backup Management	20%	{0; 0.5; 1}

The number of supported ontology languages is a factor of great importance, especially for distributed ontologies. The larger an ontology, the more important is the need for a fast querying. Ontologies can be stored in various different ways; however, poor storage can be a speed penalty under some circumstances. Therefore it is distinguished if there is intelligent and scalable underling storage. Despite all the precautions, it may happen that there is a crash and data are lost; thereby a smart backup management can avoid data loss.

Table 6.4: Further Factors to Knowledge Administration Systems.

Factor	Weight	Value Range
• Version Control	5%	{0; 1}
• Methodological Support	5%	{0; 1}
• Automatic Classification	5%	{0; 1}

The remaining three less critical, and thereby nice to have factors (see tab. 6.4), are rated on a binary scale (i.e. yes or no). A version control offers the

possibility to distinguish older from newer ontologies; the question is if the respective system offers such a version control. There are different methodologies for the creation of an ontology; the question is if a system offers such support. An automatic classification assigned something new to pre-existing classes; here again the question is if this is supported by the respective system. These three questions can be answered with yes (i.e. available) or no (i.e. not available).

All three systems support RDFS, OWL, and SPARQL that is used by the FORA framework; and in addition AllegroGraph and Jena provide both also a corresponding reasoner. Hence, AllegroGraph and Jena support four languages, KAON three. Therefore AllegroGraph and Jena are rated with 0.6 each, KAON, however, only with 0.4. Concerning the speed of retrieval all three systems have its troubles. AllegroGraph browser view must grapple with the fact that the underlying database system checks consistency and integrity and thereby forfeit its speed. Based on these reasons AllegroGraph is rated with 0.5. Jena and KAON forfeit speed as soon as the ontology no longer can be stored in the Random-Access Memory (RAM) but must be retrieved from a file. Nevertheless, both systems can be rated with 0.75. The question after the storage has just been answered; AllegroGraph stores the ontology in a graph database, whereas Jena and KAON are using traditional database (that are either not easily extensible or do not scale well). Therefore the former is rated with 1, the latters with 0. Even for the backup management the three systems can be allocated along the storage type: Jena and KAON provide an inadequate backup system that is rated with 0.5. AllegroGraph, by contrast, supports transactions with a great backup management system that is rated with 1.

The remaining three features were offered by none of the three systems in the required form. On these grounds AllegroGraph wins the race mainly because of its excellent implemented database management system. Table 6.5 illustrates the consolidation of the presented value analysis comparison of the three knowledge administration systems [Osswald, 2011].

Table 6.5: Comparison of Knowledge Administration Systems.

Factor	Weight	AllegroGraph	Jena	KAON
• Number of supported Languages	20%	0.6	0.6	0.4
• Speed of Retrieval	25%	0.5	0.75	0.75
• Type of Storage	20%	1	0	0
• Backup-Management	20%	1	0.5	0.5
• Version Control	5%	0	0	0
• Methodological Support	5%	0	0	0
• Automatic Classification	5%	0	0	0
Total	**100%**	**0.65**	**0.41**	**0.37**

With the selection of AllegroGraph as knowledge administration system now the YouReputation prototype's knowledge base can be implemented. The

following section provides a comparison of three different knowledge representation systems that are needed to point up the knowledge administration system-inherent knowledge structure in a user-understandable way.

6.3.3 KNOWLEDGE REPRESENTATION SYSTEMS

In order to evaluate the usability of knowledge representation systems for the FORA framework, also different methods of knowledge representation are compared in a value analysis. Within this PhD project, likewise no new knowledge representation system is implemented. In fact, three knowledge representation systems are selected from a range between formal and visual knowledge representation [Burkhard, 2011]. This range is deduced from [Pellegrini & Blumauer, 2006]'s semantic stair:

- *Tag Clouds:* This system provides a visual way of knowledge representation. They are based on folksonomies and offer an optical representation of these folksonomies, by visualizing different frequencies of tags. More frequent tags are represented in bigger font-sizes than less frequent ones (see fig. 4.5). Tag Clouds are not standardized and hence independent of any institution or organization.

- *Topic Maps:* This system separates information and knowledge bases to represent knowledge independently from the underlying information. Topic Maps formally describe topics, relationships and occurrences, to model knowledge contained in external information resources (see fig. 4.6). They are standardized by the International Standardization Organization (ISO) and were originally not intend for Web use. However there exists a Web-optimized standard called XML Topic Maps (XTM) that describes knowledge using XML and URI.

- *RDFS/OWL:* These systems model ontologies to represent knowledge. As already introduced in chapter 2, in the Semantic Web exist two more or less expressive ontology languages, the RDFS and OWL. In the following comparison both languages are treated as one, using the generic term RDFS/OWL. In RDFS/OWL knowledge is represented by formally modeling classes, properties and individuals. To provide interoperability, the W3C standardizes both languages. Thus they are official components of the Semantic Web.

The three selected systems are compared from the standpoint of [Herczeg, 2009]'s casual users of the Social Semantic Web. Casual users are of interest, since they only spend limited time with the very same information system. This corresponds to the observed usage of different systems through the analyzed media users (see chap. 5). Because of infrequent use, casual media users do not evolve routines. It is thereby negligible how much experience they have in using a certain information system.

Table 6.6: HCI in Knowledge Representation System.

Factor	Weight	Value Range
• Visualization	20%	{0; 0.25; 0.5; 0.75; 1}
• User Incentive	15%	{0; 0.5; 1}
• Comprehensibility for User	15%	{0; 0.5; 1}

GUI design is a practice whose techniques are encompassed by the field of HCI [Hearst, 2011]. Therefore, the first criterion of the value analysis quantifies the HCI ease (see tab. 6.6). It is subdivided into three measurable factors: The factor visualization quantifies the visual support when interacting with a system. The second factor measures the user incentive to start and keep using a certain system. The factor user comprehensibility indicates how easy the functional principle of a system can be understood.

The second criterion rates the efficiency of a knowledge representation system. This is illustrated in table 6.7. It is again divided into three factors: The first factor standard states the interoperability of a certain system, regarding the concept of the Social Semantic Web. The factor modeling method rates the efficiency on how a system is modeling knowledge. The factor critical size determines the size a system needs to achieve to be able to perform logic-based operations.

Table 6.7: Efficiency of a Knowledge Representation System.

Factor	Weight	Value Range
• Standard	12%	{0; 0.25; 0.5; 0.75; 1}
• Modeling Method	9%	{0; 0.25; 0.5; 0.75; 1}
• Critical Size	9%	{0; 0.5; 1}

The third criterion indicates the complexity of a system (see tab. 6.8). It is measured by using two factors: The first factor extensibility quantifies the support to connect knowledge bases in order to create a Web of knowledge. The factor semantic expressivity rates the capability of a system to represent and compute meanings contained in given information.

Table 6.8: Complexity of a Knowledge Representation System.

Factor	Weight	Value Range
• Extensibility	12%	{0; 0.5; 1}
• Semantic Expressiveness	8%	{0; 0.5; 1}

The following comparison of the value analysis is clearly dependent on subjective influences [Burkhard, 2011; Hearst, 2011; Manning et al., 2008]. There are several issues that make it hard to validate the usability of different systems. The main reasons arise out of the fact that users have different needs and experiences [Herczeg, 2009]. Hence each user has a specific understanding of usability and defines various requirements. Thus the obtained result depends on the explicit standpoint the evaluation is taking

place. Furthermore, the result depends on the specified criteria, factors and weights. Because there are no generic criteria on how to measure the usability of a system, the obtained results are not generally valid. Moreover, as [Hearst, 2011] mentioned, using an information system performed tasks can vary. A different task may have an impact on the usability requirements of a system. Additionally, the way on how users interact with a certain system can differ. That is why the fuzziness of a task and the experience of a user implicate different needs on the system's support of interaction. A common determination of the usability of a system is therefore only hardly possible.

However, the consolidation of all weights of all criteria and factors of knowledge representation are shown in table 6.9. The three systems were assessed together [Burkhard, 2011]; this assessment is thereby based on subjective perception. A weight can be interpreted as the relative importance of a criterion or factor for the over-all usability of a system from the standpoint of casual users. Accordingly, HCI is the most important criterion of the analysis. The factor visualization has thereby a higher weight than the other factors user incentives and comprehensibility. For casual media users Topic Maps boast the most appealing visualization, close followed by Tag Clouds. RDFS/OWL for these users fails. Hence, Topic Maps are rated with 0.75, Tag Clouds with 0.5 and RDFS/OWL with 0. Also the user incentive is highest for Topic Maps, followed by Tag Clouds and RDFS/OWL placed a distant third. Therefore Topic Maps gets 1, Tag Clouds 0.5 and RDFS/OWL again 0. Comprehensibility yields the same picture.

Efficiency is the second most important criterion for casual users. The factor standard has the highest weight, followed by the two factors modeling method and critical size. Following a Web standard RDFS/OWL is a clear winner, followed by Topic Maps that are standardized by ISO (i.e. ISO 13250). Only Tag Clouds are not standardized. Because of that RDFS/OWL is rated with 1, Topic Maps with 0.5 and Tag Clouds with 0. For the factor modeling method the casual users get the most from Topic Maps, followed from Tag Clouds. Hence, Topic Maps are rated with 1, Tag Clouds with 0.75. RDFS/OWL is too complicated and earns 0, again. For the critical size, this order does not change since Topic Maps can be used relatively straightforward almost similar to Tag Clouds. RDFS/OWL shows a bigger critical size. Therefore Topic Maps are rated with 1, Tag Clouds with 0.5 and RDFS/OWL with 0.

The lowest importance is assigned to the criterion complexity, whereas the factor extensibility has a higher weight than the factor semantic expressivity. However, here clearly both standardized systems are ahead by a nose. Since the FORA framework is concerned with the Social Semantic Web, RDFS/OWL is a bit more appropriate than Topic Maps. Hence both factors extensibility and semantic expressiveness are rated same: RDFS/OWL with 1, Topic Maps with 0.5, and Tag Clouds with 0.

As table 6.9 indicates, have Topic Maps the highest total value and hence provide the best usability in the considered context. Casual users' major requirements regarding the usability of a system are an easy and immediately understandable HCI. Topic Maps fulfill these requirements by supporting the easiest way of knowledge representation in the considered comparison. Its functional principle and modeling method are based on the fuzzy human cognition and are therefore simple to understand. Moreover the possibility to visualize Topic Maps by drawing semantic networks provides an excellent way to search and explore information needed. Unfortunately there is no generally accepted visual representation defined in the Topic Maps standard.

Table 6.9: Comparison of Knowledge Representation Systems.

Factor	Weight	RDFS OWL	Topic Maps	Tag Clouds
• Visualization	20%	0	0.75	0.5
• User Incentive	15%	0	1	0.5
• Comprehensibility for User	15%	0	1	0.5
• Standard	12%	1	0.5	0
• Modeling Method	9%	0	1	0.75
• Critical Size	9%	0	1	0.5
• Extensibility	12%	1	0.5	0
• Semantic Expressiveness	8%	1	0.5	0
Total	**100%**	**0.32**	**0.79**	**0.36**

For the implementation of the YouReputation prototype, a proprietary knowledge representation widget is used to visualize topics and tags using interactive Topic Maps. In doing so largely the example of Topic Maps Martian Notation (TMMN) is followed [Pitts, 2009]. Based on TMMN, the findability of information is improved. It constitutes a simple graphical notation used to map out ontologies and representative instances. Related tags are displayed using interactive Topic Maps, enabling a communications operative to find related tags by browsing. Similar topics (i.e. ontologies) and appropriate tags (i.e. representative instances) are visualized closer and the more dissimilar topics and tags are placed farther apart (see chap. 4). The topic contains a set of related tags presented on the screen and allows the clicking of any tag that appears around the topic.

Comparable to [Zadeh, 2010]'s z-mouse, the dashboard allows the communication operatives to zoom in-and-out (akin to the zooming function in Google Maps[34]) to find related topics and associated tags for a stated query as [Hearst, 2011] proposes. Hence, this interactive visualization helps to identify the previously unknown but related topics and tags and to thereby gain new knowledge.

Please note that for Topic Maps dedicated engine were developed. A Topic Maps engine is a knowledge administration system based on the Topic Map ISO 13250 standard exposing a Topic Map API compatible interface for running Topic Map applications. Nevertheless, the ISO 13250 standard is so far not merged into the W3C standard. Since this PhD project is concerned with fuzziness in the Social Semantic Web, it was basically built upon W3C's Social Semantic Web standards if possible. Therefore in this section a comparison of Topic Maps engines is waived.

6.4 IMPLICATIONS FOR THE PROTOTYPE

As good travel maps help navigate reliably through areas on route, conceptual frameworks are designed to do the same within the development of design science artifacts. The first emerging artifact is the introduced FORA framework. This framework consists of a modular-constructed architecture of components to support communication operatives in their daily business.

To support the framework, the main components (i.e. algorithms and systems) were assessed through comparisons such as value and performance analysis. Yet, for the implementation of the FORA framework (i.e. as proof of concept), a separation of concerns between information acquisition and the use of information in systems is taken in hand as a guiding principle:

- *The information acquisition*: Is concerned with the research foci of artificial intelligence (i.e. Web agents help find appropriate information from Web), aggregation of information (i.e. fuzzy clustering algorithms helps to put the appropriate information in context), context (i.e. Web agents and the ensuing aggregation of information help to understand the information context), and storage (i.e. disk-based graph database to help manage found information). The process of information acquisition mostly deals with an algorithmic view and should be extensible and scalable as another stated research focus.

- *The use of acquired information*: Deals with different possible approaches presenting computer-found information and let users naturally interacting with this information. The foci covered are user interaction (i.e. users manipulate presented information to better understand it), knowledge representation and reasoning (i.e. through an automatic information representation a user can derive new knowledge), and context interpretation (i.e. users can perceive computer-found context information).

As a result, abstraction is derived that helps acquire, collect, and manage information in a system-independent fashion and identify corresponding software components. These modular and loosely coupled components form the basis of the following second artifact of this design research—the YouReputation prototype implementation.

6.5 FURTHER READINGS

Following a design sciences research approach as described by [Hevner et al., 2004; Simon, 1996], this section introduced first the testable technical requirements for the FORA framework. [Hay, 2002] detailed illustrates these requirement analysis techniques. [Miles & Huberman, 1994] and [Maxwell, 2005] are helpful writings for conceptual frameworks. These frameworks are used in research to outline possible courses of action or to present a preferred approach to an idea or tough—as the FORA framework.

Parts of the FORA framework are inter alia mentioned in [Portmann, 2009; Portmann & Meier, 2010; Portmann, 2011a; Portmann & Kuhn, 2010]. Thereby some articles are concerned more with technical aspects of information aggregation, whereby others deal with knowledge representation. The ideas to visualize the context part of the dashboard of the FORA framework have been developed together with [Hächler, 2010]. However, in this PhD thesis, the basic considerations were massively extended. The complete FORA framework is described in [Portmann et al., 2012]. The comparisons of the different best practices for the FORA framework components were created with the help of several Bachelor students at the University of Fribourg. Their thesis can be found at [Burkhard, 2011; Kolly, 2011; Osswald, 2011]. A more complete list of the by [Osswald, 2011] compared different ontology storage systems can be found by [Portmann et al., 2012], where also ontology query languages are compared.

7
THE YOUREPUTATION PROTOTYPE

"A learning machine is any device
whose actions are influenced by past experiences."

—Nils Nilsson

The rise of the Social Semantic Web brought with it the always connected way of life. As conversations are increasingly distributed, everything starts with listening and observing the conversationscape. Doing so will help organizations to identify exactly where relevant discussions are taking place, as well as their scale and frequency. The YouReputation prototype is the entrepreneurial venture to bring different aspects of the presented FORA framework to fruition. It is an algorithmic and systematic instantiation of the framework's interaction and navigation, achieved through a separation of information acquisition and the use of acquired information. Hence, YouReputation constitutes an information design prototype, distinct from both the look and feel of final software.

The FORA framework is the first output of the chosen design science research approach. However, this first output needs to be validated, thereby producing a second research output—the YouReputation prototype. YouReputation is a portmanteau formed by contracting the word your with the word reputation, voicing in this way the importance of individual online reputation management. In the course of this, [Lim et al., 2008] suggest that, by prototyping, a design idea can be implemented for the purpose of evaluation. Hence, the proof of the pudding is the eating. Following this, the YouReputation prototype is a tangible piece of (free) Web-based software that admits experiencing the key ideas of the FORA framework without programming it entirely. The prototype is used to instantiate a possible application of the framework, as well as to illustrate a promising smart interaction with social media elements. At that, two different key aspects are considered: Socio-semantic Social Web requirements stemming from communication operatives (i.e. for the use of acquired information) and technical re-

quirements arising from state of the art of the Semantic Web technology (i.e. for information acquisition).

To back communication operatives' effort in listening and observing the conversationscape, a computer can consult the Social Semantic Web that accounts for the YouReputation prototype data. As a method of unsupervised machine learning, fuzzy clustering allows a computer in its information acquisition process to learn to recognize complex patterns in data by default. That is why the YouReputation prototype can be considered as a learning machine. However, for the creation of the fuzzy grassroots ontology, the lex parsimoniae is followed as described in [Portmann, 2011a]. This law of parsimony suggests tending towards simpler solutions until some simplicity can be traded for increased explanatory power. This law can be applied to all situations that require a more efficient, functional solution. When resources are limited or when speed of function is essential, design and complexity trade-offs should be based on what does the least harm to the probability of success, however that is defined. This law likewise follows suit for the implementation of the knowledge representation as well. So that the communication operatives can perceive (i.e. use) the computer-acquired information, it has to be visualized interactively. By the design of the two widgets (i.e. knowledge representation and context-based hit list) simplicity was a key goal with which strict observance helped avoiding unnecessary complexity. The implementation's underlying approach was to keep the prototype implementation simple.

On that account section 7.1 first demonstrates the challenges of the FORA framework and the consequent emerging kernel of the YouReputation prototype. Section 7.2 reveals information acquisition as the first part of the prototype. In section 7.3 the second part of the prototype is illustrated: the use of the acquired information. In section 7.4 an evaluation of the YouReputation prototype is performed. There the distinction is made between the information acquisition and its usage. In the former, cluster quality criteria are evaluated, whereby in the latter corresponding communication operatives' requirements are evaluated. In addition, a comparison to comparable applications on the market is undertaken. To help get the gist quickly and fed back into reality, section 7.5 highlights the major points of the FORA framework and the YouReputation prototype (i.e. design science research artifacts). Finally, section 7.6 presents the most important further readings.

7.1 Introduction to the Prototype

The FORA framework (see chap. 6) permits searching the Social Web (see chap. 2) to find reliable information on reputation (see chap. 4). Using this framework, it is possible to scan social media elements according to a query to determine topic classes with related tags and, thus, to identify hidden information. The determination of the topic classes is achieved by means of fuzzy clustering (see chap. 3). For the implementation of the YouReputation prototype (i.e. as instantiation of the framework) the law of parsimony is followed. This goes hand in hand with the applied methodology of prototyping. To effectuate the communication operatives' requirements from chapter 5 for the YouReputation prototype, structured techniques are used to design a final system.

The development process started with the evolution of preliminary models and its instantiations. These pre-alpha stage models consisted of individual and independent parts of software. In the next stage, the requirements are verified using prototyping, whereby the pre-alpha models are updated to a first alpha version of the YouReputation prototype. This alpha version ignited the software testing phase that continued to be repeated iteratively. As a consequence the YouReputation prototype constitutes a free Web-based service that is located between alpha and beta stage of (Web) software development. Such being the case, the prototype still exhibits certain instability but, after all, it can be tested by (Web) users. This beta release is, however, useful for the demonstration of the PhD project underlying concepts. The selected (rapid) prototyping approach therefore entails compromises in functionality and performance in exchange for enabling faster development and facilitating application maintenance.

The creation of the individual parts of the YouReputation prototype originates from a co-work of the University of Fribourg's Information Systems[35] research group with the Keio-NUS CUTE[36] center. The group Information Systems at the University of Fribourg (Switzerland) focuses among others on information retrieval, fuzzy classifications and mediamatics (i.e. media and informatics). Thereby technologies, strategies, and methodologies from multimedia are combined with social media to simplify HCI and at that initiate enhanced collective intelligence. The Keio-NUS CUTE center, however, is operated by the National University of Singapore (NUS)[37] together with Japanese Keio University[38]. Through the center, researchers of the two universities collaborate on research themes such as lifestyle media in the ubiquitous society and global computing, while utilizing leading-edge network and trends in digital content and Asian pop culture. During the authors research stay at the center in summer 2010, in a small team first parts of the YouReputation prototype were evaluated and implemented. These ideas and the corresponding prototype were afterwards cultivated, mainly at the Information Systems research group.

The prototype is provided as Web service (i.e. as API that supports interoperable computer-to-computer interaction) on a Drupal[39] website—retrievable under the address: `http://youreputation.org`. To support the editing of identified issues (see chap. 5), the Web API is defined as a set of HTTP request messages along with a definition of the structure of response messages. The website runs on a platform that supports both a Web server capable of running PHP and a database to store content and settings. The website's content is thereby stored in a traditional database that is deployed with Drupal. Thereby HTML was used for the definition of the structure and Cascading Style Sheets (CSS) for the layout of the website. The fuzzy grassroots ontology, however, is as the prototype's content stored to the introduced AllegroGraph RDFStore (see chap. 6) that is queried using SPARQL. Besides, the prototype uses PHP on the server side to access AllegroGraph, and to query Delicious as well as Twitter. On the client side JavaScript is employed to interactively visualize the Topic Map. Thereby two JavaScript libraries were used: Raphaël for Web vector graphics and jQuery for scripting HTML. Both libraries are straightforward and cross-browser compatible. Lastly, JSON is used for data exchange between PHP and JavaScript.

The underlying fuzzy grassroots ontology was created with Matlab that allows matrix manipulations (i.e. text input), plotting of data (i.e. compiling a tagspace), implementation of algorithms (i.e. fuzzy clustering algorithm), as well as interfacing with third programs (i.e. AllegroGraph; see chap. 6). Matlab's input was crawled from Delicious Web service on March 9, 2011 and stored in text as matrix format. Using [Bourke, 1997]'s algorithm, this matrix data were plotted on a tagspace and then processed with [Fu & Medico, 2007]'s FLAME algorithm to build the fuzzy grassroots ontology (see chap. 6). Figure 7.1 illustrates the YouReputation prototype kernel. This kernel is separated in two parts: One concerning the information acquisition, the other the usage of information. The information acquisition part comprises the following subparts:

- *Tag slurp:* The tag slurp is designed as a Web agent that travels across the Social Web to collect metadata (i.e. tags from folksonomies). In the current release the tag slurp is based on one-shot data from Delicious.

- *Tag purifier:* The heterogeneous tags must be normalized before processing. This is the task of the tag purifier that clears the (tag slurp received) tags and passes it afterwards on to the tagspace creator.

- *Tagspace creator:* The tagspace creator interrelates the purified tags and plots them on a plane (i.e. the tagspace). The relationship is calculated using a proximity measure and the plotting is based on a calculation of intersection points between circles.

- *Ontology adaptor:* Then, as the last processing step, the ontology adaptor converts the tagspace to the fuzzy grassroots ontology using fuzzy clus-

tering methods. The fuzzy grassroots ontology is stored in OWL format that can be queried with SPARQL.

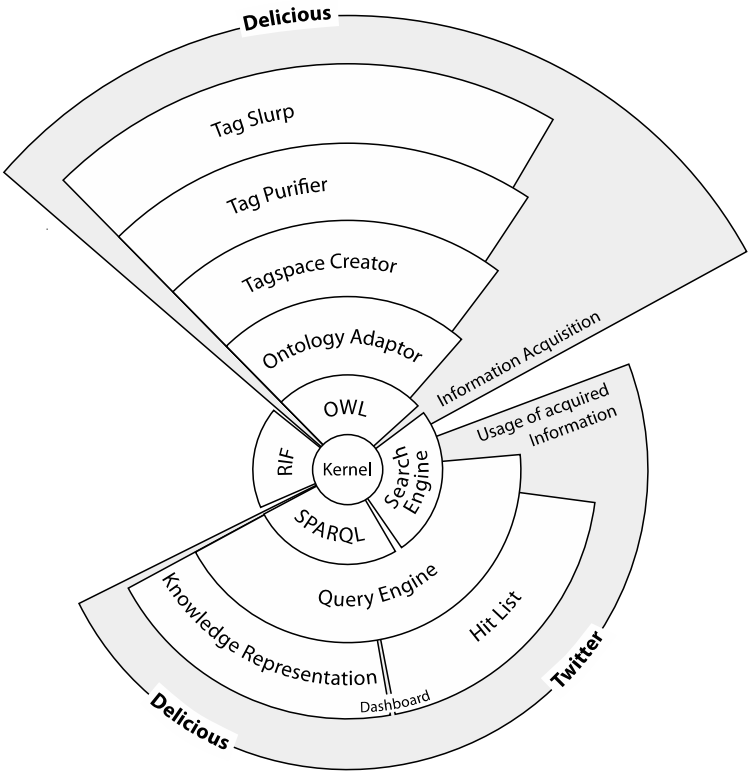

Figure 7.1: YouReputation Prototype Kernel.

The goal of the YouReputation prototype is to provide useful social media information concerning a user's reputation. Therefore it answers a query with appropriate information from the Social Web. Note that since the YouReputation prototype is not commercially used and goes as proof of concept for a fuzzy online reputation analysis application, here no single communication operatives (i.e. as a representative for an organization) is addressed but all Web users instead. Using fuzziness, human and computer can approach each other. With fuzzy clustering techniques, vague human perceptions (i.e. used for perceiving the world and expressed in folksonomies) can be converted to computer-understandable, accurate models (i.e. ontologies); this may lead to enhanced collective intelligence. This is the usage part of the previously found information that comprises a query engine, and a dashboard applet containing, in turn, a knowledge representation and a context-based hit list widget:

- *Query engine:* The query engine translates user queries to SPARQL in order to query the fuzzy grassroots ontology. Afterwards it sends the ontology-enhanced search to Delicious (i.e. as status dimension) and Twit-

ter (i.e. as time dimension) and collects their hits which are then shown on the dashboard applet.

- *Knowledge representation widget*: This widget displays the user query term matched part of the underlying fuzzy grassroots ontology as Topic Map. This matched part of the ontology can be expanded or restricted by a respective analogous zoom function (i.e. to experience the underlying ontology's fuzziness).

- *Context-based hit list*: This widget displays the query engine collected hits containing links to (Social) Web sources that correspond to the users query. Through clicking on a link a user can visit the source and hence interact in an appropriate manner.

In the YouReputation prototype not all elements of the FORA framework (see chap. 6) are implemented. To date, the search engine queries Delicious and Twitter. Because of that, so far, no metasearch engine approach is coming into operation. However, this search engine will be enhanced to a metasearch engine through adding more underlying Web service (i.e. as proposed in chap. 5 and 6).

Another so far unexploited part is the possibility to include rules into the underlying knowledge administration system in which also the fuzzy grassroots ontology is stored. Using RIF, for example, rules derived from sentiment analysis systems could be introduced into the prototype as well. From a human-backed sentiment analysis as presented by [Go et al., 2009], for example, vague rules could be derived (by default). In interaction with a fuzzy reasoning engine [Simou & Kollias, 2007] over a fuzzy extended DL language [Stoilos et al., 2005a; Bobillo & Straccia, 2008] the fuzzy grassroots ontology could incorporate sentiment analysis rules as well.

Among other things, its modular structure is the YouReputation prototype's strong point. The prototype is extensible and scalable with Semantic Web technology. If and when Semantic Web evolves, this can be included into the FORA framework and hence also in the YouReputation prototype. Whenever the user inputs a search, the query engine attempts to find relevant terms by querying the underlying knowledge base.

7.2 Information Acquisition

This first information acquisition part, of the prototype constitutes the learning element of YouReputation. Its unsupervised classification is a method of artificial intelligence and a key task in machine learning, by which the prototype learns to automatically recognize patterns, to discriminate between elements on the basis of their variant patterns, and to make reasonable decisions (see chap. 3).

In the following section 7.2.1, the tag slurp is presented. In the second part, the found tags are normalized and the underlying sources are ranked. To that end, section 7.2.2 introduces the tag purifier that normalizes the collected tags before they are processed further. After the tags have been collected and normalized, they need to be sorted. Section 7.2.3 demonstrates the normalizing and plotting of the found tags, named tagspace creator. After this step, all of the tags are connected to each other and plotted onto a tagspace. The fourth phase is the ontology adaption, which separates the plotted tagspace into hierarchies of classes. This is presented in section 7.2.4. The ontology adaptor splits the plotted tags into meaningful categories, thereby evoking a fuzzy grassroots ontology.

7.2.1 TAG SLURP

At the beginning, Web agents identify all n tags with the underlying sources by constantly crawling the Social Web, and subjoin them into lists (i.e. as text file format). For each tag a text file is generated that contains all tag-related tags (incl. their frequency of joint appearance). These in comma-separated text files stored collected tags from folksonomies are needed to establish the fuzzy grassroots ontology. A corresponding customizable Web agent to harness folksonomies was created during a Bachelor project. However, [Oggier, 2009]'s Web agent is intended mainly for demonstrating purpose, that is, to display one possibility to gather user-created tags in a semi-automated fashion. There are typically two ways of acquiring data: First using an API, which has the advantage that it is quite easy to implement. However, if data from multiple different pages should be acquired then the use of different APIs becomes necessary. The second way is to download whole webpages and extract its content. This has the advantage that it also works with pages which do not provide an API and that the Web agent can be written in a generic way, such that it can be configured for different webpages easily.

Since the continuous updating leads to an extremely fast-growing data volume (see chap. 5) that cannot be addressed adequate with the available financial, technical, and computational resources (i.e. hardware and software), for the YouReputation prototype only once a record of Delicious data was obtained. To save limited resources, in addition, of these tags only the most frequently jointly used tags were further processed. In order to fetch the Delicious tags, a two-stage crawler was developed: In the first stage, the crawler fetched all the tags existing in Delicious with their corresponding appearance. To simplify the prototype, only the top 500-appearance tags are taken into consideration. In the second stage, the crawler collected the data from the combination of each two tags in the top 500 tags set. Finally, the co-appearance of each tag with another high-appeared tag is saved in a text file and also in memory. Note that, in order to collect such data via two stages, the duration for multi-threaded crawler is six days.

However, for the underlying knowledge base, the ability to find high-quality sources (i.e. tagged sources in future social media elements) is important for overcoming information overload. Collaborative filtering or recommender systems can identify high-quality sources that utilize individual knowledge. Several ranking algorithms use link-based centrality metrics, including Google's PageRank and Kleinberg's Hyperlink-Induced Topic Search (HITS) algorithm. In [Baeza-Yates et al., 2011] these two algorithms are presented as the most widespread; of these, in [Liechti, 2012]'s Bachelor thesis, Google's PageRank is evaluated as a more appropriate for the FORA framework. A more elaborated ranking scheme consists of using linear combination of different relevance signals. For example, relevance signals that in the case of online reputation analysis could be of importance are log file, page tag or sentimental analysis of the connotation of a tag (see chap. 4). This could be included as rules from a respective system using RIF. As described by [Stoilos et al., 2006], different research is going on to include fuzzy rule languages into the Semantic Web that aim to capture and handle different types of uncertainties.

Because the data were obtained only once from Delicious, at this point, no link analysis was performed. In a next improvement phases, the continuous updating as well as the link analysis would have to be considered and implemented to obtain an improvement in the data corpus. Thereby an upgrade from constantly crawled Delicious tags to further constantly crawled folksonomies (e.g. from Twitter, Facebook, etc.) is conceivable. Though, these additional tags can cause processing disorders since only for their storage far more space must be made available. Besides the approximate calculations of the ontology needs, in that case, also more time to complete. However, in return it would be reflected in an improved dashboard applet, and reputation search in general.

7.2.2 TAG PURIFIER

As is generally known, folksonomies can contain erroneous tags. According to [Lewandowski, 2005], different error correcting strategies have to be distinguished. Dictionary-based approaches compare tags with a dictionary and if the dictionary does not cover the tag, they search for similar ones. Statistical methods refer misspellings with no or only a few hits to the most commonly used similar syntax. As a result, to correct faulty tags, several work steps must be completed. The first step is to transform all characters in the text files to lowercase. With the use of a phonetic algorithm, yet another step follows. In the course of these work steps, however, linguistic issues can be resolved: First, homographs are detected by comparing all tags character strings. Homophones, in turn, are recognized using a phonetic algorithm. To determine phonetic similarity, tags are reduced to a code that is able to conform to similar tags. Since synonyms show a high similarity, they are detected during the tagspace creation as last linguistic issue.

The phonetic Metaphone algorithm, transforms consonants to codes. Thereby also digraphs are involved and accordingly transformed. A digraph is a pair of characters used to write one phoneme or a sequence of phonemes that does not correspond to the normal values of the two characters combined. The most common used digraph in order of frequency in the English language is *'th'*, for example. Vowels are also used, but only at the beginning of the code. The standard procedure of the basic Metaphone algorithm is shown in algorithm 7.1.

Algorithm 7.1: Metaphone Algorithm.

1. Drop duplicate adjacent letters other than *'c'*.
2. If the word begins with *'kn'*, *'gn'*, *'pn'*, *'ae'*, or *'wr'*, then drop its first letter.
3. Drop *'b'* if it is after *'m'* and if it is at the end of the tag.
4. Transform *'c'*:
    a. If it is followed by *'ia'* or *'h'* to *'x'* (unless *'h'*, it is part of *'sch'*, in which case it is transformed to *'k'*).
    b. If followed by *'i'*, *'e'*, or *'y'* to *'s'*
    c. Else to *'k'*.
5. Transform *'d'*:
    a. If it is followed by *'ge'*, *'gy'*, or *'gi'* to *'j'*.
    b. Else to *'t'*.
6. Drop *'g'*:
    a. If it is followed by *'h'* (unless *'h'* is at the end or before a vowel).
    b. If it is followed by *'n'* or *'ned'* and is at the end.
7. Transform *'g'*:
    a. If it is before *'i'*, *'e'*, or *'y'*, and it is not in *'gg'* to *'j'*.
    b. Else to *'k'*.
8. Drop *'h'* if it is after and not before a vowel.
9. Transform *'ck'* to *'k'*.
10. Transform *'ph'* to *'f'*.
11. Transform *'q'* to *'k'*.
12. Transform *'s'*:
    a. If it is followed by *'h'*, *'io'*, or *'ia'* to *'x'*.
13. Transform *'t'*:
    a. If it is followed by *'ia'* or *'io'* to *'x'*.
    b. If it is followed by *'h'* to 0.
    c. If it is followed by *'ch'* drop *'t'*.
14. Transform *'v'* to *'f'*.
15. Transform *'wh'*:
    a. If it is at the beginning to *'w'*.
    b. If it is not followed by a vowel then drop *'w'*.
16. Transform *'x'*:
    a. If it is at the beginning to *'s'*.
    b. Else to *'ks'*.
17. Drop *'y'* if it is not followed by a vowel.
18. Transform *'z'* to *'s'*.
19. Drop all vowels unless it is the beginning.

This illustrates the simplest form of the Metaphone algorithm that was presented for reasons of comprehensibility. However, meanwhile, there are extensions of this algorithm such as the Double Metaphone as well as Metaphone 3 [Phillips, 2000]. These algorithms use a much more complex rule set to account for further irregularities. Different realizations of these extended algorithms are currently in place in various (Web) programming languages (e.g. PHP, JavaScript, etc.) and do not have to be implemented from beginning. Such an algorithm was also used for the correction of the tagspace. A major advantage is that the correctly spelled ontology terms could be used as auto-completion and auto-suggestion while the user is typing search terms into the dashboard for example. A log study by [Anick & Kantamneni, 2008] found that users clicked on such dynamic suggestions about one third of the time they were present. Yet, this is not implemented into the YouReputation prototype so far. Neither is there as yet an implementation of the matching of synonyms with a lexical database. In a next step, however, the inclusion of such a database (e.g. WordNet[40]) is planned.

In the next improvement phase of the prototype, also a stemming of the tags could be applied ex ante to further improve the fuzzy grassroots ontology-underlying data corpus. Using a stemming algorithm, inflected (or derived) tags (e.g. plurals, gerund forms, past tense suffixes, etc.) can be reduced to their stem to circumvent syntactical variations that prevent a perfect match. As mentioned, within the YouReputation prototype so far no stemming is performed that can—together with the possibly too simple stop word list—lead to messy topics in the dashboard-inherent Topic Map. There are a variety of stop word lists but there is not one standard list. In a next phase, though, a more suitable list could be selected.

7.2.3 TAGSPACE CREATOR

After all of the tags have been collected and normalized, they need to be sorted. The tagspace is a two-dimensional plotted representation of a consistent picture and serves as the input for the ontology adaption. Several steps are required to plot the tagspace from the found tags. The first step is to define the relationship of the various found tags. According to [Hassan-Montero & Herrero-Solana, 2006], the simplest way to define these relationships is to use the Jaccard similarity coefficient $d_J(A, B)$, presented in chapter 3. Therein relative co-occurrence is identical to the partition among the amount of resources in which tags co-occur and the amount of resources in which either of the two tags appear. This can be calculated using the normalized text files (see sect. 7.2.1 and 7.2.2). Besides, this method causes tags to become united and offers a semantically consistent picture in which nearly all of the tags are related to each other. This semantically consistent picture is referred to as tagspace (i.e. a distance matrix). Based on a similarity coefficient also synonyms can be detected, as introduced in the previous section.

Now, to begin plotting the point representation of the tagspace, it is necessary to set a limitation for the tagspace. The plotting algorithm starts with a number of seed points as illustrated by algorithm 7.2. Some seed points will be referred from the seeds, but they are limited to a certain depth with the distance d being a stated distance (see chap. 3). Child point locations are computed based on [Bourke, 1997]'s classical algorithm, which calculates the intersection of two or three circles.

Algorithm 7.2: Plotting Points.

1. Create a tag list from a number of seeds with a predefined depth and select one source tag.
2. Select each tag in the list except the selected tag.
3. Calculate the plotted tags that are within a given distance d to the selected tag.
4. Check the number of plotted tags that have a relationship with the current tag:
    a. If no plotted tags are detected, then draw the current tag with a random position.
    b. If there is one plotted tag detected, then draw the current tag with the same y but with a x value that is calculated to fit the distance.
    c. If there are two plotted tags detected, then draw the current tag as one of the two intersections point of two circles whose centroids and radii are the two plotted tags and their distances to the current tag, respectively.
    d. If there are three plotted tags detected, then draw the current tag as the intersection of the three circles whose centroids and radii are the three plotted tags and their distances to the current tag, respectively.
5. Return to Step 2 for the next point.

After the found and normalized tags have been united, assorted and plotted onto a tagspace, a computer-understandable ontology can be established. The algorithm allocates the position of each point in the tagspace. Based on this algorithm, the necessary points in the selected region can easily be shown, which is very effective for supporting a zoom function. Another parameter to take into account is the constant variability of the underlying data. Normally data are at fixed values to be analyzed, but here, they are constantly moving around. In fact, they change every week, hour or second (depending on the Web agents update frequency). This consideration is legitimate because data come from real world, where no absolutes exist. The trends or demands of the Web can change acute. To interact with live data, they need to be continually updated. As a result, the introduced plotting algorithm is able to provide a good perspective on moving data [Portmann et al., 2012].

7.2.4 Ontology Adaptor

The ontology adaption can be described as follows: All n tags plotted on the tagspace will be sorted by [Fu & Medico, 2007]'s FLAME algorithm. This algorithm is based on an approximation of neighbor tags on the tagspace. To this end, the algorithm starts with the identification of Cluster Supporting Objects (CSO). These CSOs constitute representative tags (i.e. prototypes)

around which the cluster will be constructed. To do this, the similarities between each pair of tags and the k-Nearest Neighbor (KNN) are selected to weigh the density $\rho = 1/d\,;\ \forall x$ around the representative tags. For this purpose each tag is connected with its KNN, with k being a predefined constant. The FLAME algorithm as an outline is presented in algorithm 7.3.

Algorithm 7.3: FLAME Algorithm.

1. Extract the tagspace-inherent structure by constructing a neighborhood graph where each tag x_j is connected to its KNNs.
2. Calculate a density ρ for each tag based on its distance d to its nearest neighbor using the density formula $\rho = 1/\mathrm{d}$.
3. Assign each tag:
    a. If the tag has a higher density ρ than all its neighbor tags to a cluster supporting object $C_{CSO} = \{C_1, \dots, C_{c-1}\}$.
    b. If the tag has a density lowers than all its neighbors, and lowers than a threshold ε (i.e. an arbitrary constant) to cluster outliers $C_O (= C_c)$.
4. Assign the tags of the cluster supporting objects C_{CSO} full membership to its respective clusters $\Gamma_{CSO} = \{\Gamma_1, \dots, \Gamma_{c-1}\}$.
5. Assign the tags of the cluster outliers C_O full membership to its respective cluster Γ_O.
6. Assign for every remaining tag membership to all clusters C (incl. the outlier C_O).
7. Reiterate assigning membership until the approximation error $E(\{u_{ij}\})$ converges to zero by updating the memberships u_{ij} of the remaining tags by a linear combination of the memberships of its nearest neighbors (assign higher weights w_{ij} for closer objects x_j).

In Algorithm 7.3 the weights $w_{ij} \in [0,1]$ define how much each neighbor will contribute to approximation of the fuzzy membership of that neighbor; the sum of all weightings is 1. In the course of this, the approximation error is:

$$E(\{u_{ij}\}) = \sum\nolimits_{x_i \in X} \left\| u(x_i) - \sum\nolimits_{x_j \in KNN(x_i)} w_{ij} u(x_j) \right\|^2 .$$

However, all of the clusters are stored using a knowledge administration system, so several clusters that are jointly called fuzzy grassroots ontology are obtained. Following [Bobillo & Straccia, 2011]'s approach, a procedure to represent the ontology-inherent information within current standard languages and tools is strived for. The created fuzzy grassroots ontology now only needs to be stored using OWL [Bobillo & Straccia, 2011; Stoilos et al., 2005b]. The returned websites (i.e. ABox) that belong to the single tags are ranked and stored separately but linked to the ontology (i.e. TBox) to establish a knowledge base. The entire knowledge base is stored using AllegroGraph's RDFStore (see chap. 6).

7.3 USAGE OF ACQUIRED INFORMATION

This second part of the YouReputation prototype is mainly concerned with the usage of the acquired (and transformed) information. It constitutes the knowledge representation elements of YouReputation. The focus is on understandability by humans that supports recognizing social media context. Thereby the computer-learned human ontologies are returned to them. This supports enhanced collective intelligence since computers learned something from humans (i.e. domain classification through folksonomies) what they now pass back graphically (i.e. as Topic Maps and hit lists separated into context dimensions). This graphical return is visualized on the dashboard applet. This applet is an interactive GUI designed that its visualization is easily read; it is the part of the prototype that users (e.g. communication operatives) interact with. Besides the applet, a further equally important part of the system is the query engine, with which automatically presented queries are created after first use. Every user interaction on the dashboard-visualized Topic Map prompts the query engine to provide a new SPARQL query to find related topics and tags within the fuzzy grassroots ontology. Once the related topics and tags have been located, the query engine also provides the dashboard with the stored and ranked underlying websites. To this end, the query engine mediates between the dashboard and the underling fuzzy grassroots ontology.

In the following two sections the dashboard applet will be introduced in broad. This applet constitutes a JavaScript application that provides the interactivity that cannot be provided by HTML alone. In response to the users action these dashboard applet change the provided graphic content. In doing so, the applet consists of two widgets: Section 7.3.1 presents the knowledge representation widget that visualizes knowledge by Topic Maps in a user understandable manner, and section 7.3.2 illustrates the hit list widget that displays the found hits according its context dimensions. The hits are coming from Delicious (i.e. status dimension) and Twitter (i.e. time dimension). Last but not least, in section 7.3.3 the query engine is illustrated.

7.3.1 KNOWLEDGE REPRESENTATION WIDGET

The dashboard applet is the main visualization of the system. It provides a knowledge representation widget that conveys information such as topics and tags and the relationships between them (see chap. 4 et seq.). The visualization not only shows Topic Maps that were inducted from search results but also more valuable information, such as the different layers (multi-level) that can be viewed by zooming in. The Topic Map helps to identify search results by topics that users (e.g. communications operatives) can focus on to find exactly what they are looking for or to discover unexpected relationships between items. Remember that tags visualized farther away from a topic belong to it at a less significant level than do the tags that are closer; the same applies to the relationship of the topic itself. Nevertheless, each time the user manipulates a search weight K via a slider (see fig. 7.2) or

clicks on a topic (i.e. in blue) the missing parts are inserted into the dashboard applet.

The first time the dashboard applet is used, users do not need to adjust any settings (e.g. search weight K) but only feed the search box with a name, product, or brand. The search weight $K \in [0,1]$ denotes the expressed broad of the search defined by the underlying fuzzy grassroots ontology. Thereby $K = 0$ states that simply the term in the search box is searched (i.e. Boolean search), and $K = 1$ the contrary (i.e. every related term in the underlying ontology is searched). This search box was implemented in accordance with a casual user's appetence for simplicity [Herczeg, 2009]. The interactive visualization should intuitively lead the user to his desired Web contents. Based on the query engine, the dashboard provides a suggested indicator. In other words, K needs not to be set manually but is automatically set by the framework (i.e. $K \approx 0.3$). However, a user may change the weight (e.g. to $K \approx 0.7$) by dragging the slider on the left side of the knowledge representation widget. Using click and zoom functions, a user can evaluate an entered search term on the knowledge representation widget of the dashboard applet (i.e. the user can adjust K implicitly).

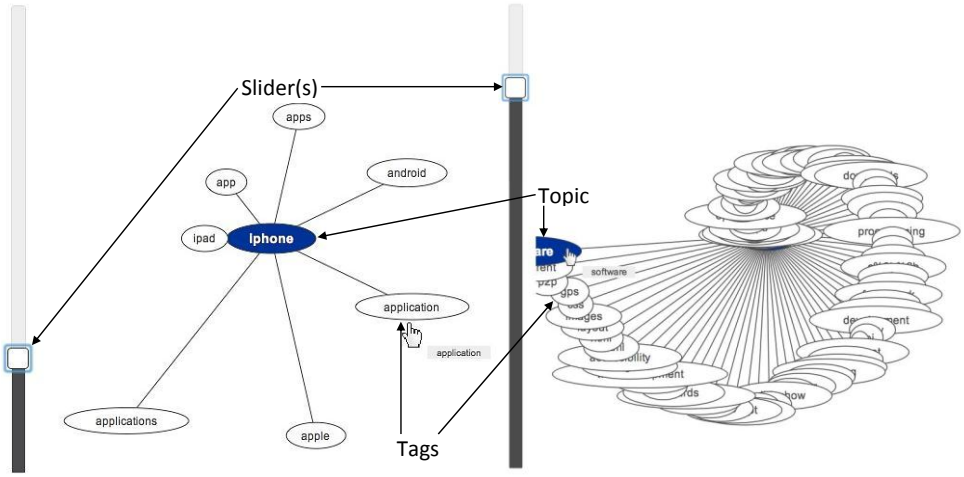

Figure 7.2: The YouReputation Knowledge Representation Widget.

In figure 7.2 on the left the (in ascending order sorted) Topic Map of an initial search for iPhone is illustrated (i.e. as helix). As mentioned, thereby the YouReputation prototype selects K by default. Now a user can drag the slider and select a new K. The slider thereby shows the underlying fuzzy concept by being metonymic with an analog world-perception (i.e. in contrast to a digital one that rather would be illustrated by an exact digit). Nevertheless, figure 7.2 right-hand illustrates the changed Topic Map after the users alteration of K. By clicking any of the tags or topics around the iPhone topic, the search can be restricted (i.e. by clicking on the topic related tags such as ap-

plication or gps) or expanded (i.e. by clicking on a new topic such as software).

The fuzzy search is always based on the user entered search term. Therefore this term is placed in the center of the knowledge representation widget. However, based on K, surrounding topics (i.e. from the underlying fuzzy grassroots ontology) are displayed on the Topic Map. By clicking these surrounding topics (i.e. software in blue) these topics move to the center with their own tags (i.e. restricted by the same K as well). Furthermore, the visual displays hits in different context dimensions, allowing gaining further knowledge not only about the entered search term. This is introduced in the next section.

7.3.2 HIT LIST WIDGET

For Web search engines, users are given to look at only the top-ranked retrieved results and are biased toward thinking the top results are better than those beneath it, simply by the SERP position [Hearst, 2011]. To stand up to this, a smart representation of the topic-corresponding hits can support further insights. A good way to present hits are [Dey & Abowd, 2000; Dey et al., 2001] minimal context dimensions to display. The different characteristics of social media can help to distinguish these minimal context dimensions; some are better in achieving such a distinction, others less (see chap. 5 and 6). Splitting hits with respect to their origin into context dimensions, allows an intuitive interaction with different kind of social media.

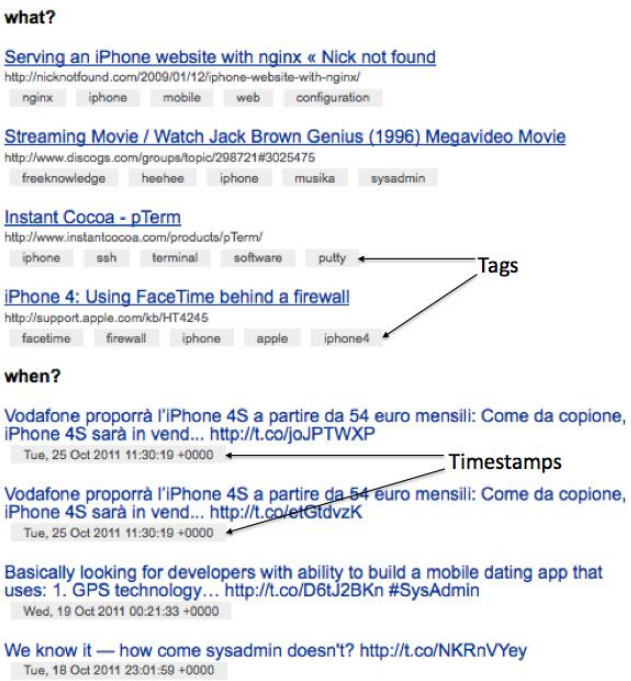

Figure 7.3: The YouReputation Hit List Widget.

As YouReputation goes as proof of concept, only two social media elements are implemented. First, Delicious for the status dimension and second, Twitter for the time dimension. Both of them are integrated via their APIs. On one hand, using Delicious users' collective intelligence, this Web service allows searching for specific results and thereby also presents to each found hit respective user-assigned tags (i.e. what?). This can be abstracted from figure 7.3. So, under each found hit there are additional user tags as further suggestion to search. Through clicking these tags, the search can be further restricted (i.e. not yet implemented).

On the other hand, Twitter is a great tool for finding temporal information (i.e. when?). Their real-time search allows a chronological distinction of found results. This is illustrated at the bottom of figure 7.3. Here, under each found hit there are vague temporal tags (i.e. timestamps). Through clicking these timestamps, the search can be further restricted too (i.e. also not implemented). Note that by clicking on every tag—no matter whether the tags on the Topic Map or under the respective hits found—every time the hit list changes and includes real-time hits.

In any case, presenting found hits according minimal context dimensions are a first step to ease HCI. As a result a user gets a sense of the underlying Web data and its importance. Besides, using fuzziness can intensify the ease of interaction. At the moment, the order of the hit list is specified by the underlying Web services (i.e. Delicious and Twitter). However, this order could also be determined by sentiment analysis rules introduced into the prototype using RIF. Furthermore, using different colors to visualize the importance of found sources (see chap. 5) can be supported by fuzziness as well. For instance for sentiment analysis tasks returning an analog scale (e.g. by color shades) rather than a digital judgment, correlation is a better measure than precision because it takes into account how close the predicted value is to the target value. Each hit could be visualized by a color mixing of green shades (i.e. more or less positive mentions), yellow shades (i.e. more or less neutral mentions), and red (i.e. more or less negative mentions) for example. As a result, each hit would be presented in its individual color on which basis the impact could be optically measured. So in general the dashboard applet could support users in forecasting process (see chap. 4). For example, the monitoring of a topic over time may provide indication for trends as presented in figure 4.9. The not yet implemented color scale, however, could support forecasting process as well.

7.3.3 Query Engine

The query engine transforms user requests to SPARQL, and simultaneously communicates (via their APIs) with the Delicious and Twitter search engine. It is an introduction into how a user-provided query can be enriched using the fuzzy grassroots ontology. So, after building the fuzzy grassroots ontology, it feels perfectly natural to be up to performing so-called conjunctive

queries, which take the form $\exists x_1 \ldots x_n(q_1 \wedge \ldots \wedge q_n)$, where $q_2, \ldots, q_n$ are further search terms. Classical (i.e. hard) DL encounters the bothersome issue that if a tuple does not satisfy the strict constraints of a users' query, then it is not included in the result (i.e. Boolean search). This happens because the membership or non-membership of a term to certain topic is built using the law of excluded middle that says $\Pi \vee \neg\Pi$ (see chap. 3).

In contrast, the fuzzy grassroots ontology provides a solution by which the deficiencies can be hurdled. Instead of specifying the exact membership degrees that a term (or pair of terms) should belong to a topic, the constraints can be left unspecified the first time. Now, whenever the system receives search input from a user, the query engine performs semantic queries to find the terms that are near to the input term regarding its semantic.

Initially, SPARQL returns all terms that are at a distance closer than approximately 0.3 (the preset weight for K) to the search term iPhone. The prototype specifies the weight by default with a membership degree using SPARQLs `FILTER` function (see chap. 2). The result would be to retrieve every tuple of the knowledge base that participates in the assertions to a special degree (e.g. $K \approx 0.3 \Rightarrow$`?distance<0.3`). In addition, the sequence modifier `ORDER` helps putting the results in ascending order (see helix form in fig. 7.2). Then inferring the conjunctions as fuzzy intersections, a ranking of tuples can be provided and presented the user with an initial set of the most relevant information (e.g. `ipad+iphone`). To do this, the found terms (i.e. `ipad+iphone`) are now searched in the underlying two Web services (i.e. Delicious and Twitter) by availing their API. Afterwards, by interacting with the Topic Map, the user can verify if more results should be fetched.

Since SPARQL is at the moment limited (see chap. 2), please note that, according to [Stoilos et al., 2005b], answering such types of queries accordingly is still an open research issue. Yet, [Cheng et al., 2010] present a flexible extension of SPARQL. They illustrate how to efficiently compute the top answers of flexible SPARQL queries with regard to membership degrees and user-defined weights (e.g. as suggested here by interactive Topic Maps). However, this is so far not yet been incorporated into the prototype that way.

In the end a word on the requirements of an editing functions from chapter 5. Unlike the dashboard applet inbuilt knowledge representation widget (i.e. to react to social media mentions) and the hit list widget (i.e. to put social media mentions in context), the editing function cannot be selected directly on the YouReputation start page. Rather the prototype is based on the W3C recommendation of Web Services Description Language (WSDL) 2.0 specification [Booth & Liu, 2007; Chinnici et al., 2007]. This specification offers support for RESTful Web API. Hence, the YouReputation prototype's Web API adheres to [Fielding, 2000]'s principles of REpresentational State Transfer (REST) architecture that and can be accessed straightforward through:

`http://youreputation.org/resources/`. This API can be used to download found issues as spreadsheet, to prompt trigger functions from a third system, and to integrate found issues into such a third system.

7.4 EVALUATION

The YouReputation prototype provides proof of the FORA framework. However, in this section, in turn, the YouReputation prototype itself is evaluated. Since within the design strictly the law of parsimony is followed, the result is a prototype closer to a mockup than a ready product. However, manufacturing a prototype requires YouReputation to be in a way complete for the users handling. Yet, the YouReputation prototype is built on the Semantic Web Stack and thereby is extensible. Hence, the prototype can prosper along with the evolution of the Semantic Web.

To keep consistency, the evaluation of the prototype is, again, separated into information acquisition and the use of information. Representative for the acquired information, in section 7.4.1, the fuzzy grassroots ontology is evaluated as final product of the learning element of YouReputation. The evaluation is performed together with [Kolly, 2011]. The same procedure is followed for the evaluation of the use of information. For that purpose in section 7.4.2 the dashboard applet is compared with products on the market, based on the communication operatives requirements. This comparison is performed at PostFinance Inc. together with [Uhlmann, 2011].

7.4.1 FUZZY GRASSROOTS ONTOLOGY

Since the fuzzy grassroots ontology depends on clustering techniques, it is obvious to apply its validation techniques (see chap. 3). The most used clustering validity methods include the evaluation of the tradeoff between cluster compactness, separability and stability. There are no benchmark data to compare the FLAME algorithm obtained cluster quality. However, from a human perspective (i.e. expert aspect), the quality of the clustering results looks well. Yet, since in future the fuzzy grassroots ontology gets updated automatically and, as the name unsupervised learning task implies, thus can barely be benchmarked automatically, at this point an external criterion is abstained from.

Objective functions in clustering formalize the purpose of attaining high intra-cluster similarity and low inter-cluster similarity (i.e. internal criterion). As shown in chapter 3, for a good clustering (i.e. with high information entropy) I_D should be as high as possible and I_{DB} as low as possible. In contrast to the former, the latter index is based on distances that are easily adapted to fuzzy clustering. In [Kolly, 2011]'s Bachelor thesis, the $I_{DB}$is adjusted to $\frac{1}{c}\sum_{i=1}^{c} max_{i\neq j}\left(\frac{\sigma_i+\sigma_j}{d(c_i,c_j)}\right)$, with $\sigma_i = \frac{1}{n}\sum_{a=1}^{n} d(x_a, c_i)u_i(x_a)$ that implies conformity with hard but marks out also as fuzzy David-Bouldin index I_{fDB}. In

the course of this, each term has full membership to the cluster in which the term lies (i.e. $u_i(x_a) = 1$), and 0 for the others.

For comparison, Sandro Kolly iteratively calculated various I_{fDB} for FCM, GK, and FLAME algorithms with a randomly generated test set. In doing so, FLAME always turned out to be the algorithm with the lowest I_{fDB}. For the crawled data from Delicious Web service the I_{fDB} amounts to 0.5048. Awkwardly there are no comparative data, but because this value is below 1, this outcome is to consider as quite well. This hypothesis thereby is supported by [Kolly, 2011]'s test sets.

However, a good outcome on an internal criterion does not automatically translate into good effectiveness in an application. According to [Manning et al., 2008] an option to internal criteria is the direct evaluation of the application of interest. This relative criterion is examined with the evaluation of the YouReputation dashboard in the next section.

7.4.2 Dashboard Applet

According to [Hearst, 2011], to evaluate a prototype's GUI, subjective measures are equally or even more important than quantitative measures, because if a user has a choice between two systems he will use the one he prefers (e.g. for aesthetics, familiarity, preferred features, perceived ranking accuracy or speed reasons). How best to evaluate the dashboard applet depends on the stage in the development process. Because the YouReputation prototype is still located between alpha and beta version, the applet therefore is evaluated using a discount usability method [Nielsen, 1989]. This is a methodology for cheap usability evaluation that includes narrowed-down prototypes, simplified user testing, and heuristic evaluation. Often it yields better results than deluxe usability because it drives an emphasis on early and rapid iteration with frequent usability input. Besides, this method goes hand in hand with the chosen approach of rapid prototyping.

In chapter 5, the communication operatives' requirements to an online reputation analysis application have been assessed. For those purpose communication operatives at Swiss financial services institutions were interviewed together with [Uhlmann, 2011]. Then the interviewed communication operatives' opinions were evaluated, spawning a set of requirements for online reputation analysis applications (see chap. 5). However, following [Nielsen, 1989]'s discount usability method, also the YouReputation prototype is assessed. In doing so, together with communication operatives at PostFinance, [Uhlmann, 2011] qualitatively evaluated the dashboard applet. To do this, the applets design functionalities were evaluated in terms of how well it measures up against comparable applications on the market to the communication operatives' requirements from chapter 5. The compared applications thereby are:

- *Sysomos[41] heartbeat:* This is a well-structured but relatively expensive application. Sysomos is headquartered in Toronto (Canada) and offers apart from heartbeat further social media analysis applications. Yet, in comparisons, the heartbeat application is regularly top-ranked [Schwede & Stöcklin, 2011]. It mainly impresses by its design, interactivity, user friendliness and graphical presentation of hits.

- *MeMo News[42]:* This is the product from the leading Swiss provider of monitoring applications. The application is constantly extended and, in the course of this, adapted to different social media elements. MeMo news offers its customers individually tailored solutions and is characterized by a superior individual support.

- *Actionly[43]:* Actionly is privately held and is based out of San Francisco (USA) and Mumbai (India). It allows a monitoring of different social media elements and, in doing so, a management directly out of the application. It works fully automated and without any personal support.

Table 7.1 illustrates the applications performance in comparison to the YouReputation prototype. A more complete comparison can be found in [Uhlmann, 2011].

Table 7.1: Comparison of Selected Applications.

Factor		Sysomos	MeMo	Actionly	YouReputation
1.	**Usability**				
	Look & Feel	Excellent	Poorly	Good	Simple
	Forward & Share	Yes	No	Yes	API
	Respond	Yes	No	Yes	Yes
	Data Export	Spreadsheet	No	Spreadsheet	Spreadsheet
2.	**Languages**				
	Language Choice	DE, EN, FR, IT	DE, EN, FR, IT	DE, EN, FR, IT	EN
3.	**Quantitative data analysis**				
	Graphics	Excellent	Good	Good	Good
	Duplicates & Spam	Yes	Yes	No	Yes
4.	**Qualitative data analysis**				
	Keywords	Yes	No	Yes	Yes
	Key Topics	Unreliable	No	Unreliable	Yes
5.	**Reporting**				
	Alert Functions	Email	No	Email	API
	Overview by Email	Daily	Daily	Daily	No

In contrast to the YouReputation prototype, Sysomos heartbeat, MeMo news and Actionly provide all broad implemented functions. On these grounds, up next follows a rough-and-ready summary of [Uhlmann, 2011]'s evaluation of YouReputation's dashboard applet. To this end, in the next paragraph the positive findings are presented.

Instead of endless and often confusing hit lists the YouReputation's dashboard applet provides relevant contributions in clear manner and thereby meets the requirements of communication operatives for reasonable manual expenditure and final qualitative reports. In contrast to other applications, the search for opinion leaders, new and relevant topics and important keywords is no longer necessary. Yet, the YouReputation prototype recognizes the correlations of the keyword by its knowledge base, and arranges similar results to bundles. In this way, the online reputation analysis requirement to put mentions in context is complied. Moreover, the YouReputation prototype is ahead of others in terms of supporting semantic analysis. However, it must be said that the prototype works so far only for certain search terms and new keywords have to be learned first. Through the implementation of a continuous automatic update (see sect. 7.2.1), the prototype would learn the keywords fully automated. Due to the PhD projects limited financial, technical, and computational resources this implementation was omitted in the first step. On the other hand, a performance testing was not entirely omitted but since the YouReputation prototype is still between alpha and beta stage (see sect. 7.1) only a simple user experience under load test was run. Thereby, because a common run is not expected any time soon, only normal and no peak load conditions were tested.

For the test the reviewers used the dashboard from different locations roughly the same time. Thereby each reviewer independently entered unspecified search terms and stopped the time to load the Topic Map and the dedicated hit list. On average (i.e. out of 30 trials), it takes 1.05 seconds to load the Topic Map (shortest loading time was 0.9 seconds, the longest 1.4 seconds), and 3.86 seconds to load the hit lists (shortest loading time was 2.2 seconds, the longest 4.8 second). Because the hit list widget depends on the knowledge representation widget, it makes sense that the load of the hit list is slower. Notwithstanding there is still development potential, for the YouReputation prototype these loading times were regarded as good enough.

Table 7.2: Alignment of Online Reputation Analysis Requirements.

Requirement	Comment
• React to Mentions	This is only partially complied because it is not possible to respond with mentions within the application.
• Put mentions in Context	This is mainly complied with the application.
• Edit Found Mentions	This is not complied because it is not possible to edit mentions directly within the application but an API must be used.

Table 7.2 summarizes the findings in relation to the elaborated key requirements of chapter 5 [Uhlmann, 2011]. In conclusion it can be said that the YouReputation prototype is an appropriate application for reputation analysis, but the following core functionalities are missed, without which the application will not be used in financial services institution: If there is a de-

mand to respond immediately to issues, alert functionalities (i.e. not via API), regular summaries (i.e. via e-mail) and a simple forwarding (i.e. also not via API) are very important. In comparison to other applications, communications operatives can partially intervene with the analysis by hand. Thus, contributions cannot be directly analyzed manually (e.g. by date, source, importance, etc.). Hence, the reaction to mentions is complied with a limited extend. So, the edit of found mentions requirements has up to this point potential for improvement. However, the YouReputation prototype's objective is that these tasks are done by the application itself. If and when the YouReputation prototype becomes ready for the market, for communication operatives this would be a promising future prospect.

In summary, it can be stated that none of the applications on the market meet the requirements of communications operatives at PostFinance in full. All conventional applications must expand their semantic functionalities towards YouReputation, and YouReputation, in turn, its user-friendliness and reliability towards the products on the market.

7.5 SYNOPSIS

On the basis of the FORA framework, in this last chapter the YouReputation prototype is presented. Since its underlying field of online reputation management adjoins to diverse paradigms of research (e.g. cognitive science, communication, information systems, linguistics, marketing, mathematics, media, and logic), this PhD project is built upon three different yet related research fields:

- *Social sciences*: Commonly used as a collective term to refer to fields outside of the natural sciences such as communication, linguistics, marketing, media, and others. Here, the main research focus is on integrated reputation management.

- *Information systems*: An applied academic field at the hinges of social and natural sciences that is based on the main on business and computer science. Within this PhD thesis the focus of information systems research is essentially on HCIR.

- *Computer sciences*: A natural and applied sciences related field concerned with information and computation foundations, as well as implementation and application practices. It often intersects other disciplines, such as cognitive science, linguistics, mathematics, logic, and others. At this point, the emphasis is on the Social Semantic Web.

Figure 7.4 illustrates this PhD project's backing of three research fields with each with its most focused-on domain. In the intersection of all these fields, in turn, emerge the FORA framework and the YouReputation prototype.

The four-way intersection of Social Semantic Web with HCIR uncovers Social Web and HCI; the one of Social Semantic Web with integrated reputation management the Semantic Web and online reputation analysis. Last, the intersection of HCIR with integrated reputation management reveals IR and context dimensions of social media elements. The emphasis is, however, on fuzziness because this concept is so powerful that all these research fields can be linked together. For one thing fuzzy logic is a unifying concept rooted in real world (and not in models) and, for another thing, it is able to translate vague social science concepts to hard applied and natural science models and back, and thereby eases the tension between humans and computers.

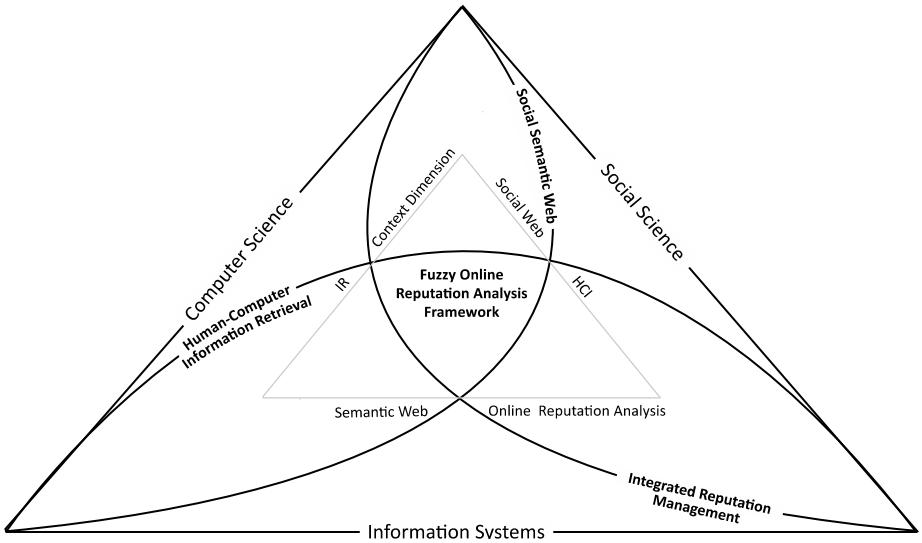

Figure 7.4: FORA Framework Background Triangle.

However described, all these intersections are influencing the FORA framework. In addition, for the implementation of the YouReputation prototype, a separation of concerns between information acquisition and the use of information in systems is taken in hand as a guiding principle. As a result, abstraction is derived that help acquire, collect, and manage information in a system-independent fashion and identify corresponding software components. These modular and loosely coupled components, along with a simple distributed platform, form the basis of the prototype. Hence, all seven research foci from chapter 5 (i.e. user interaction, knowledge representation and reasoning, context interpretation, artificial intelligence, aggregation of information, storage, and extensibility and scalability) are assembled. With the validation of the YouReputation prototype, which goes as proof of concept of the FORA framework, things have come full circle.

7.6 Further Readings

[Nielsen, 1989; Snyder, 2003] introduces paper prototyping as the fastest and easiest way to design and refine user interfaces and [Chua et al., 2010] present remarkably rapid prototyping methods. A highly recommended introduction into effective prototyping for software makers is provided by [Arnowitz et al., 2006] as well as [Bernard & Summers, 2010], which induct into dynamic prototyping. In [Hearst, 2011] and [Lim et al., 2008], prototyping particularly for the fields of HCI, software engineering, and design is envisaged as a specific kind of object used in the design process.

The FORA framework is first outlined by [Portmann et al., 2012]. Parts of the framework are sketched in [Portmann, 2009; Portmann & Kuhn, 2010; Portmann & Meier, 2010] and [Portmann, 2011a]. Different students participated with their Master or Bachelor projects to the development of the prototype: [Oggier, 2009] introduces a Web agent that constantly crawls the Social Web for folksonomies, [Liechti, 2012] presents a comparison of ranking algorithms, [Kolly, 2011] evaluated and implemented the FLAME clustering algorithm, [Osswald, 2011] the knowledge administration system, and [Burkhard, 2011] the knowledge representation system. [Uhlmann, 2011] evaluated the prototype together with respective communication operatives at PostFinance Inc. For the evaluation, [Nielsen, 1989]'s discount usability method was followed.

CONCLUSION

"Simple things should be simple,
complex things should be possible."

—Alan Kay

Solving problems in place, on an individual, organizational or even global level comes down to methodical access to knowledge. Networks are so vital in order to enable developing new and innovative solutions that cultivate and disseminate collective knowledge. Networks of people and organizations must necessarily be aligned with networks of knowledge. While knowledge is generally highly cross-linked, at times, this cross-linking of Social Web data is still hard to see. As a consequence, today's social media elements can prove impractical for the expansion of new and innovative solutions. So the lack of this cross-linking can hinder elementary information management and problem-solving potentials, such as finding, creating and deploying the right knowledge at the right time. Accordingly, a semantic extension of the Social Web is highly aimed at since among social media elements sometimes only little knowledge is exchanged. Unfortunately, precisely this sparse knowledge exchange can lead to redundant knowledge bases, which in turn affects the problem of information overload.

Now the marriage of Semantic Web technologies with social media elements can lead to a global cross-linked knowledge infrastructure as anticipated by visionaries like Vannevar Bush, Doug Engelbart, Ted Nelson or even Tim Berners-Lee [Ebersbach et al., 2010; Breslin et al., 2009]. Largely, their visionary thoughts remained a utopia already for too long, as the necessary foundational technologies were not yet invented. Meanwhile, however, the ingredients for the realization of their visions seem to emerge through the Social Semantic Web. While the Social Web broadly enables sharing knowledge, the Semantic Web adds ways for interoperating across domains.

The biggest problem, however, is that computer and humans do not use the same language. Computers in the Semantic Web need a clear and precise formulation, whereas humans in the Social Web avail oneself of a subjective and imprecise language. Since natural language entails a very complicated structure that serves as basis not only for communication, but also for thinking and perceiving the world, fuzziness seems to be a good unifying element to open up humans to computers. Fuzzy logic captures human vagueness and expresses it with appropriate mathematical tools [Meier et al., 2008].

Prior to the advent of the Social Semantic Web, organizations needed only to worry about something negative when it appeared in a newspaper or on the evening news. This has changed through the automated cross-linking of social media elements in the Social Semantic Web. Since users increasingly generate more and more information and computers now can understand this information, they can help with the analysis of the information through their quick processing. Thereby much of this information is a description of perceptions expressed in natural language [Zadeh, 2006]. Today computer analysis can also include consumer perceptions (e.g. opinions), which heavily drive an organization's reputation. According to [Beal & Strauss, 2008], approximately a third of the population has reviewed something online, and anyone of those reviews can now be read by millions of people.

Exactly this global cross-linking can be analyzed with the FORA framework and its YouReputation prototype implementation. Same as in reality, consumer perceptions are imprecise, and natural language that is expressive of these perceptions is also imprecise, particularly in semantics. In the FORA framework imprecision is indicative of fuzziness, the framework and prototype's underlying technique. Using research methods lent from information systems, computer and social sciences, the requirements of media users towards the framework and the prototype were recorded. The new Web is a great research tool to discover what people are concerned with. With this in mind, the following concluding chapter first summarizes the key themes developed in this PhD project in section 8.1. Thereafter, section 8.2 dwells on future research to improve on during the PhD project started studies dealing with including vague human knowledge in the Social Semantic Web. Last but not least, section 8.3 presents an outlook of a possible future for online reputation management and Web search engines in general.

8.1 SUMMARY

The creation of this PhD project followed a design science research approach. Thereby design can be distinguished from design research by the intellectual risk, that is, the number of unknowns in the proposed design which, when fruitfully hurdled, provides a fresh innovation that makes the effort research and ensures its value. This PhD thesis contribution to science is twofold: It consists for one thing of the FORA framework, and for another thing of the YouReputation prototype. With this in mind, section 8.1.1 summarizes the main results of this PhD project, and afterwards section 8.1.2 answers the initially proposed research questions.

8.1.1 SUMMARY OF THIS PHD PROJECT

These days online reputation management stands out as an application of ever-increasing importance because of the progressive use of the Social Semantic Web. In chapter 4, online reputation management is outlined as a mission of monitoring, addressing, and rectifying mentions in the Social (Semantic) Web. The concerned organization's communication operative(s) track actions and opinions, report and react to them, and thus create a feedback loop. He watches over the whole spectrum of social media elements (e.g. blogs, microblogs, folksonomies, wikis, and social networks) for online mentions (e.g. an organization name, brands, services, products and executives characters). Thereby his focus is more than just fixing on negative mentions and setting things right again. The contrary should be the case. Through his activities he may enter into an online conversation that hopefully leads to an improved attitude towards customers that, in turn, can confirm a (positive) reputation of an organization. Through partaking in the online dialogue, he can position the organization as an opinion leader and realize new opportunities. In addition, he can systematically record information from social media that can aid in prevention of further damaging events or a crisis.

Online reputation management should bring online conversations regarding an organization into focus. To get these discussions right and thus respond properly to online reputation issues, the coherences of the Social Semantic Web should be understood. Chapter 2 addresses these coherences that can roughly be divided into two parts: The Social Web that supports and fosters social interaction and the Semantic Web that provides the technical foundations for an intelligent information processing. Both parts advanced independently and autonomously from the first-generation hypertext system known as the Web. However, a common lament about semantic technologies is, where does the Semantic Web's knowledge infrastructure come from. This is to say that both the Semantic Web and the Social Web must first have an accepted and well-used cross-linked knowledge infrastructure as a base. The Social Web contributes the data, the Semantic Web the techniques to globally cross-link these data in a computer-understandable way.

Towards this end, chapter 3 presents fuzzy clustering that arranges data objects as not hard (i.e. all-or-nothing) but fuzzy (i.e. the objects holds a membership degree that ranges between all-or-nothing). Fuzzy logic originated from fuzzy set theory—an extension of the classical notion of set—that is applied to clustering. Hence, in clustering, no predefined clusters are given but data objects group themselves so that distances between a pair of objects within the group is relatively small and those between different groups relatively large. Using fuzzy clustering it is possible to turn tags (i.e. non-hierarchical keywords assigned to Web documents) into completely automated computer-understandable ontologies (i.e. sets of concepts and the relationships between those concepts). The advantage over hard clustering is that fuzzy clustering allows tags to belong to more than one cluster with possibly different membership degrees. This allows a more natural allocation of the tags to an ontology. Hence, the FORA framework's subjacent fuzzy grassroots ontology uses this technique to discover associations between items (e.g. documents, entities, and between entities and documents).

The FORA framework is a proposition to overcome the ascertained obstacles of online reputation analysis. It focuses on the two reputation management parts of identification and analysis of reputation issues and is thereby concerned with scanning, monitoring and forecasting of issues. To this end, this thesis answers the question of how media users (e.g. employees, communication operatives or journalists) deal with online (reputation) issues. To obtain a holistic image of this process, chapter 5 presents a scenario and three case studies which analyze journalistic Web searches, employees' use of social media elements, and communication operatives' online reputation analysis. Yet, these communication operatives' most important requirement for online reputation analysis is the ability to react to social media mentions. For that to happen, found mentions should be placed in context to support their searching process. To this end the fuzzy grassroots ontology helps, together with an automatic sentiment analysis, to broaden communication operatives' horizon and to support identifying upcoming threads. Other important requirements are storing, triggering and reporting of found issues.

Chapter 6 presents the FORA framework for the analysis of online reputation issues. Because it consists of various modular building blocks it is possible to address various online reputation analysis requirements. These on technical feasibility tested requirements and related building blocks in turn add up to the FORA framework. By testing the requirements different solution possibilities were compared to each other. The heart is the fuzzy grassroots ontology, a fuzzy clustering-based, hands-off generated ontology. To set this up, Web agents recurringly seek after tags that will be normalized and, by fuzzy clustering, transformed into an ontology. AllegroGraph, a carefully selected ontology administration tool, manages the fuzzy grassroots ontology. Finally, with the help of Topic Maps for communication operatives, the fuzzy grassroots ontology-inherent knowledge is interactively visual-

ized. Thus, it is possible to recognize correlations of Web documents and, as a result, get one's head around the organization online mentions.

With a minor focus on generating the fuzzy grassroots ontology and a major focus on representing the fuzzy grassroots ontology-inherent knowledge on a dashboard, in chapter 7 a possible realization of the FORA framework is introduced. Following the law of parsimony, the YouReputation prototype is a simplified but operative implementation as proof of concept. Admittedly, it is not yet ready for market, but instead serves to highlight the principle ideas of the FORA framework. In addition, it provides users (e.g. communication operatives) with a free online tool to identify their organization's reputation: but what really makes the prototype unique is what happens after the search. Instead of purveying millions of search results into a long list, it clusters similar results together into topics. The fuzzy grassroots ontology-inherent knowledge representation with interactive Topic Maps help users explore search results by topic so they can zero in on exactly what they are looking for and, in the course of this, discover unexpected relationships between items. This helps find results that otherwise would have been missed or been buried deep on a SERP. Hence, the framework combined with this prototype represents the design research's innovation that, in the first place, substantiates the research effort and ensures its value.

8.1.2 Alignment with Research Issues

Following a design sciences research approach, the focus of this PhD project was to combine well-grounded academic research with a practice-oriented application. This section now answers the research questions stated in the beginning of this PhD thesis (see chap. 1):

1. *Evolution of the Web and appearance of fuzziness*: With an in-depth literature review the Web's evolution is explored. Thereby chapter 2 identifies the most important stages: The Web 1.0 where Web content is manually annotated by experts, followed by both the Social Web where a community annotates subject matters manually and the Semantic Web where experts specify structures to let computers automatically annotate content. The merging of these two latter evolutionary steps yields the Social Semantic Web, where the collectivity of Web users specify structures to let computers automatically annotate Web content. Also based on literature review chapter 3 presents the appearance of fuzziness that can ease the artificial perception of the real world produced by dichotomous models. Concerning fuzzy applications to the Social Semantic Web, the literature is either theoretically-driven with few practical approaches or practical-driven with moderate integration of powerful fuzzy approaches. The power of interactive knowledge representation for average users is in literature not addressed.

2. *Structuring of Social Web data through fuzzy clustering*: Likewise based on literature review, but complemented with argumentative-deductive

analysis, chapter 3 illustrates fuzzy clustering and its possible applications to the Social Semantic Web. Based on literature review, hard and fuzzy clustering methods are distinguished. Thereby fuzziness is presented as an extension of a hard, deterministic and precise perception of the surrounding world that emerges from Aristotle's law of excluded middle. On the basis of an argumentative-deductive analysis, fuzzy clustering is applied to Social Semantic Web yielding fuzzy grassroots ontologies. These ontologies induce human vagueness into the Web fabricating a more semantic Social Web. Nevertheless, these ontologies so far root in folksonomies but can be extended to other online data (e.g. computer or human-created) in a further step.

3. *Management of information through knowledge administration systems*: With test installations of most promising literature review, the thesis compares and evaluates known knowledge administration systems. On the basis of W3C recommendations thereby three different potential systems are selected as the most promising from a list of various different systems: AllegroGraph, Jena, and KAON. These three systems are compared on critical factors as well as additional nice-to-have factors. These comparison arguments come from Social Semantic Web requirements. Thereby in chapter 6 AllegroGraph is picked as the most appropriate for the management of the fuzzy grassroots ontology in the realm of the FORA framework.

4. *Analysis of Web search engine usage*: Based on a scenario, and complemented by case studies, chapter 5 presents (professional) media workers' usage of social media elements and Web search engines. Notably, two of the three case studies are based on structured interviews. In general the presented scenario and case studies are not intended to focus on the discovery of a generalizable truth or on the search for cause-effect relationships. Rather, the focal point is exploration and description. Through a future supplementation of additional case studies, the exploration and description can be enhanced. However, these research methods are put down to testable requirements for the FORA framework.

5. *Specification of online reputation management and development of an online reputation analysis framework*: Online reputation management is examined by an extended literature review followed by scenarios and qualitative interview-based case studies. In the process, the employees and executives responsible for online reputation management and analysis are found. Thereby their tasks as well as necessary adaptations of an organization to the Social Semantic Web are described in chapter 4. In addition, chapter 5 presents a special investigation in the requirements of fuzzy online reputation analysis. This detailed investigation serves as a basis for the originated FORA framework. Started with an argumentative deductive analysis of conceptual framework through literature review, found knowledge, and test installations completed, chapter 6 pre-

sents the FORA framework as a universal framework for online reputation analysis. This framework thereby is solution-oriented. If in future new insights come forth, the framework easily can be extended or amended at any time thanks to its modular structure.

6. *Evaluation, development and validation of a prototype*: With an argumentative-deductive analysis following the literature review, and Matlab tests, three fuzzy clustering algorithms are compared based on their complexity, permanence, and adaptability. Likewise, based on literature review and followed by an argumentative-deductive analysis for the implementation of a useful GUI, three knowledge representation systems for presenting the fuzzy grassroots ontology-inherent knowledge are compared on their handling, their efficiency, and their complexity. Chapter 6 presents these comparisons and chapter 7 presents the YouReputation prototype as proof of concept. Based on literature review and argumentative-deductive analysis the fundament for its implementation is given. With prototyping YouReputation is brought into being as an instantiation of the most important parts of the FORA framework. With case studies, the framework and the prototype is evaluated in Chapter 7; thereby the benefits and limitations of the system are illustrated.

8.2 Future Research

An overarching goal of this PhD project was to bring the vision of the Social Semantic Web a step closer to reality by bridging the gap between real-world users of the Social Web and the logic-based underpinnings of the Semantic Web. To this end, the fuzzy transition of human-to-computer knowledge and back has been studied. On these grounds, with various promising cases of applications for the fuzzy grassroots ontology on one hand, and knowledge representation (and reasoning) on the other hand, a software system has been considered and implemented.

The proposed fuzzy grassroots ontology can stimulate information sharing, enable sophisticated search engines, support intelligent agents and the dissemination of data, minimize data loss or repetition, and help with the discovery of resources by enabling field-based searches. For example, a variation of the fuzzy grassroots ontology is used in a project for building intelligence at the iHomeLab. Based on location data, indoor and outdoor conditions, as well as fuzzy grassroots ontology-backed search queries, the Prometheus framework supports users with helpful recommendations and information preceding a search for context-aware data [Andrushevich et al., 2011; Portmann et al., 2010]. With the same method of fuzzy clustering, the InRiNa project supports the transformation of inverted indices of Nano-related research literature to a Nano-specialized fuzzy grassroots ontology. This fuzzy grassroots ontology helps specialists to locate related papers to a corresponding query more precisely and at a much faster pace [Wehrle & Portmann, 2012]. Last but not least, in the eGlossary project a prior comput-

er-produced fuzzy grassroots ontology helps users of an innovative new kind of glossary find further definitions related to the looked-for term [Martinez, 2010; Martinez, 2011]. All these three projects are developed further independently by the respective research institutes.

The fuzzy grassroots ontology is used for the FORA framework, which first and foremost is intended to be used for online reputation analysis. An appropriate customization to a more specialized search engine would be possible too; for example with intense HCIR challenges in the form of knowledge representation and reasoning. The knowledge representation and reasoning's fundamental goal is to represent knowledge in a manner that eases drawing conclusions. In other words it analyzes how to use symbol systems (which determines the semantic structure of an object such as proposition, question, command, concept, scenario, or a system of such objects) to represent a domain of discourse, along with functions that allow formalized reasoning about objects. Recent developments in knowledge representation have been driven by the Semantic Web, and include the development of XML-based knowledge representation languages and standards (e.g. RDF, RDFS, OWL, or XTM). However these languages so far rely largely on formal logic, which average media users typically are not willing to adopt.

A different approach is to let the users implicitly convey their knowledge, instead of availing them to use formal logic. Enabling a user to browse his and other user-provided knowledge (e.g. metadata) through an interaction possibility can lead to the point that they can learn from each other. In addition, not only the user but also the computers can be trained based on these interactions (e.g. through discrimination). Knowledge representation, interaction and reasoning utilize the human and computer enhanced collective intelligence and thereby leads to approximate world knowledge bases [Zadeh, 2004]. In the Swiss National Science Foundation (SNSF)[44] granted Approximate Reasoning Methods (ARM) project, world knowledge created in this way should be tried and tested for applicability to the Social Semantic Web [Portmann, 2011b]. Thereby not only fuzzy systems but also other fields of soft computing (i.e. collection of symbiotic computing methods that involve fuzzy logic, neurocomputing, evolutionary computing and probabilistic computing [Zadeh, 1994; Zadeh, 1998]) should be considered to introduce vague human concepts to the Web. A first step to introduce these concepts would be to integrate natural language into the Web in order to realize a genuine adaptive Social Semantic Web. The essence to do that is the concept of a generalized constraint [Zadeh, 2006]. NLP seems to require extensive knowledge about the outside world and the ability to manipulate it [Manning & Schutze, 1999]. Exactly this knowledge can be provided by the proposed world knowledge bases. In particular the proposed world knowledge bases constitute a translation of aggregated human perceptions to condensed knowledge structures of the outside world. The concept of

generalized constraint can act as footing for generalizing the widely accepted view that information is inherently statistical.

8.3 OUTLOOK

There is a long way towards an adaptive Social Semantic Web. Some of its challenges include deceit, inconsistency, uncertainty, vagueness, and vastness. Deceit means that a producer of information is intentionally misleading a consumer of information. Inconsistency comes across as logical contradiction(s) that inevitably arise during the development or combination of large ontologies. Precise concepts with uncertain values yield uncertainty. Vagueness arises from the haziness of user queries, concepts represented by content providers, matching query terms to provider terms and trying to combine different knowledge bases with overlapping but subtly different concepts. Last, vastness means that a reasoning system will have to deal with huge inputs.

Nevertheless, these problems should not be approached in isolation. In fact, various emerging technologies must be harmonized to an adaptive Web. One first step to master the introduced challenges could be to properly integrate NLP into the Web in order to achieve a genuine adaptive Social Semantic Web [Zadeh, 2009; Portmann, 2011b]. Soft computing can be used for this task but there are several other emerging technologies meant to overcome further challenges toward such a Web: the WOT model (interconnection of all types of devices through Web standards to perceive outside world), machine translation (the translation of text or speech from one natural language to another), machine vision (the recognition of objects in an image and the ability to assign properties to those objects to make them computer-readable), speech recognition (which converts spoken words to text and back), and structured storage (which does not require fixed table schemas).

Thus, based on these technologies, future communications operatives will be able to use natural language to search for online reputation. Since, by then, the stated challenges (i.e. deceit, inconsistency, uncertainty, vagueness, and vastness) would be relaxed in such an adaptive Web and it will be possible to ask questions and receive reliable, context dependent answers. In [Ibaraki & Lin, 2011], Nova Spivack goes one step further and predicts a bright future in which social and computer sciences (and its information systems) have advanced so far and fast that they enable humanity and its computers turning into symbiotically one and the same. If that happens, there is a transition from parts to a whole. As of yet, humanity is far away from this prediction, but maybe an adaptive Social Semantic Web constitutes a first step towards this augury?

REFERENCES

Aarnoudse, L., 2011. *makeITfair*. [Online] Available at: `http://makeitfair.org/makeitfair-1/the-facts/news/the-story-behind-apple` [Accessed 31 October 2011].

Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, p.716–723.

Alby, T., 2008. *Web 2.0*. 3rd ed. München: Hanser.

Allemang, D. & Hendler, J.A., 2008. *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Burlington: Morgan Kaufmann.

Andrushevich, A. et al., 2011. Prometheus Framework for Fuzzy Information Retrieval in Semantic Spaces. In *Human Computer Systems Interaction: Backgrounds and Applications*. 2nd ed. Springer.

Anick, P. & Kantamneni, R.G., 2008. A longitudinal study of real-time search assistance adoption. In *Proceedings of the 31st annual international ACM SIGIR conference on research and development in information retrieval*. New York, NY, USA, 2008.

Antoniou, G. & van Harmelen, F., 2008. *A Semantic Web Primer*. Cambridge London: MIT Press.

Arnowitz, J., Arent, M. & Berger, N., 2006. *Effective Prototyping for Software Makers*. 1st ed. San Francisco: Morgan Kaufmann.

Aula, P. & Mantere, S., 2008. *Strategic Reputation Management: Towards a Company of Good*. Psychology Pr.

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R., 2010. *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Berlin: Springer.

Baeza-Yates, R. & Ribeiro-Neto, B., 1999. *Modern Information Retrieval*. New York: Addison Wesley.

Baeza-Yates, R. & Ribeiro-Neto, B., 2011. *Modern Information Retrieval: The Concepts and Technology behind Search*. 2nd ed. New York: Addison-Wesley.

Baeza-Yates, R., Ribeiro-Neto, B. & Maarek, Y., 2011. Web Retrieval. In R. Baeza-Yates & B. Ribeiro-Neto, eds. *Modern Information Retrieval: The*

concepts and technology behind search. 2nd ed. Essex: Pearson Education Limited. pp.447-514.

Baeza-Yates, R., Ribeiro-Neto, B. & Navarro, G., 2011. Queries: Language & Properties. In R. Baeza-Yates & B. Ribeiro-Neto, eds. *Modern Information Retrieval: The concepts and technology behind search*. 2nd ed. Essex: Pearson Education Limited. pp.255-79.

Baeza-Yates, R., Ribeiro-Neto, B., Navarro, G. & Ziviani, N., 2011. Documents: Languages & Properties. In R. Baeza-Yates & B. Ribeiro-Neto, eds. *Modern Information Retrieval: The Concepts and Technology behind search*. 2nd ed. Essex: Pearson Education Limited. pp.203-54.

Banks, M., 2008. *On the Way to the Web: The Secret History of the Internet and Its Founders*. Berkeley: Apress.

Barnett, M.L., Jermier, J.M. & Lafferty, B.A., 2006. Corporate Reputation: The Definitional Landscape. *Corporate Reputation Review, 9 (1)*, pp.26-38.

Barro, S. & Marin, R., eds., 2010. *Fuzzy Logic in Medicine*. Heidelberg: Physica-Verlag.

Beal, A. & Strauss, J., 2008. *Radically Transparent: Monitoring and Managing Reputations Online*. 1st ed. John Wiley & Sons.

Becker, J., Krcmar, H. & Niehaves, B., 2009. *Wissenschaftstheorie und gestaltungsorientierte Wirtschaftsinformatik*. Heidelberg: Physica-Verlag.

Bell, G., 2009. *Building Social Web Applications*. 1st ed. Sebastopol: O'Reilly Media.

Bergmann, M., 2008. *An Introduction to Many-Valued and Fuzzy Logic*. Cambridge: Cambridge University Press.

Bernard, C. & Summers, S., 2010. *Dynamic Prototyping with SketchFlow in Expression Blend*. 1st ed. Quebec: Pearson Education Inc.

Berners-Lee, T., Hendler, J. & Lassila, O., 2001. The Semantic Web. *Scientific American Magazine*, 17 May.

Bezdek, J., 1973. *Fuzzy Mathemetics in Pattern Classification*. PhD thesis. Ithaca.

Bezdek, J.C., 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. 1st ed. New York: Plenum Press.

Bezdek, J.C., Keller, J., Krisnapuram, R. & Pal, N.R., 2008. *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. New York: Springer.

Bishop, C.M., 2007. *Pattern Recognition and Machine Learning*. Berlin: Springer.

Blumauer, A. & Pellegrini, T., 2009. *Social Semantic Web: Web 2.0 - Was nun?* Berlin: Springer.

Bobillo, F. & Straccia, U., 2008. fuzzyDL: An expressive fuzzy description logic reasoner. In *IEEE World Congress on Computational Intelligence*. Hong Kong , 2008.

Bobillo, F. & Straccia, U., 2011. Fuzzy Ontology Representation using OWL 2. *International Journal of Approximate Reasoning*, 20 May.

Bondi, A.B., 2000. Characteristics of scalability and their impact on Performance. In ACM, ed. *Proceedings of the 2nd international workshop on Software and performance*. Ontario, Canada, 2000. ACM New York.

Booth, D. & Liu, C.K., 2007. *Web Services Description Language (WSDL) Version 2.0 Part 0: Primer*. [Online] Available at: `http://www.w3.org/TR/wsdl20-primer/` [Accessed 15 October 2011].

Bourke, P., 1997. *Intersection of two circles*. [Online] Available at: `http://paulbourke.net/geometry/2circle/` [Accessed 26 August 2011].

Bray, T., Hollander, D.H. & Layman, A.L., 1999. *Namespaces in XML*. [Online] Available at: `http://www.w3.org/TR/1999/REC-xml-names-19990114/` [Accessed 3 August 2011].

Breslin, J.G., Passant, A. & Decker, S., 2009. *The Social Semantic Web*. 1st ed. Berlin: Springer.

Brickeley, D. & Guha, R., 2004. *RDF Vocabulary Description Language 1.0: RDF Schema*. [Online] Available at: `http://www.w3.org/TR/rdf-schema/` [Accessed 3 August 2011].

Bruhn, M., 2002. *Relationship Marketing: Management of Customer Relations*. Essex: Financial Times Prent.Int.

Budura, A., Michel, A., Cudré-Mauroux, P. & Aberer, K., 2009. Neighborhood-Based Tag Prediction. In *Proceedings of the 6th European Semantic Web Conference on The Semantic Web: Research and Applications*. Heraklion, Crete, Greece, 2009. Springer.

Burkhard, P., 2011. *Benutzerfreundlichkeit ausgewählter Wissensrepräsentationssysteme im Sozialen Semantischen Web*. Bachelor Thesis. Fribourg: University of Fribourg.

Burkhardt, S., 2009. *Reputationsmanagement: Ziele, Strategien und Erfolgsfaktoren*. Berlin: Bundesverband deutscher Pressesprecher Hamburg Media School.

Bush, V., 1945. As We May Think. *The Atlantic Monthly*, July.

Cardoso, J., 2007. The Semantic Web Vision: Where Are We? *Society, IEEE Computer*, pp.22-26.

Cheng, J., Ma, Z.M. & Yan, L., 2010. f-SPARQL: A Flexible Extension of SPARQL. In *Proceedings of DEXA*., 2010. Springer Verlag.

Chermack, T.J., 2011. *Scenario Planning in Organizations: How to Create, Use, and Assess Scenarios*. San Francisco: Mcgraw-Hill Professional.

Chinnici, R., Moreau, J.-J., Ryman, A. & Weerawarana, S., 2007. *Web Services Description Language (WSDL) Version 2.0 Part 1: Core Language*. [Online] Available at: `http://www.w3.org/TR/wsdl20/` [Accessed 15 October 2011].

Chua, C.K., Leong, K.F. & Lim, C.S., 2010. *Rapid Prototyping: Principles and Applications*. 3rd ed. Singapore: World Scientific Publishing Company.

Chun, R., 2005. Corporate Reputation: Meaning and Measurement. *International Journal of Management Reviews, 7(2)*, pp.91-109.

Comm, J., 2009. *Twitter Power*. Hoboken: John Wiley & Sons.

de Oliveira, J.V. & Pedrycz, W., 2007. *Advances in Fuzzy Clustering and its Applications*. West Sussex: John Wiley & Sons Ltd.

Dey, A.K. & Abowd, G.D., 2000. Towards a better Understanding of Context and Context-Awareness. In *CHI 2000 Workshop on the What, Who, Where, When, and How of Context-Awareness*. Hague, 2000.

Dey, A.K., Abowd, G.D. & Salber, D., 2001. A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. *Human-Computer Interaction*, pp.97-166.

Donges, P., 2008. *Medialisierung politischer Organisationen: Parteien in der Mediengesellschaft*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Doorley, J. & Garcia, H.F., 2010. *Reputation Management: The Key to Successful Public Relations and Corporate Communication*. New York: Routledge Chapman & Hall.

Dover, D., 2011. *Search Engine Optimization Secrets*. Indianapolis: John Wiley & Sons.

Dunn, J., 1974. A fuzzy relative of the isodata process and its use in detecting compact, well separated clusters. *Journal of Cybernetics*, pp.95-104.

Eberl, M., 2006. *Unternehmensreputation und Kaufverhalten: Methodische Aspekte komplexer Strukturmodelle*. Wiesbaden: Deutscher Universitäts-Verlag | GWV Fachverlage GmbH.

Ebersbach, A., Glaser, M. & Heigl, R., 2010. *Social Web*. UVK Verlagsgesellschaft mbH.

Ebert, T., 2009. *Trust as the Key to Loyalty in Business-to-Consumer*. Wiesbaden: Gabler Edition Wissenschaft.

Eccles, R.G., Newquist, S.C. & Schatz, R., 2007. Reputation and Its Risks. *Harvard Business Review*, 01 February. pp.104-18.

Eck, K., 2010. *Transparent und glaubwürdig: Das optimale Online Reputation Management für Unternehmen*. München: Redline Verlag.

Eisenegger, M., 2008. Issue Monitoring: Früherkennung und Analyse öffentlicher Kommunikations- und Reputationsdynamiken. *fög discussion papers*, February.

Eisenegger, M. & Imhof, K., 2007. *fög discussion papers*. [Online] University of Zurich Available at: `http://www.foeg.uzh.ch/staging/userfiles/file/Deutsch/`

f%C3%B6g%20discussion%20papers/2007-0001_True_Good_Beautiful_e.pdf [Accessed 4 August 2011].

Eisenegger, M. & Imhof, K., 2009. Funktionale, soziale und expressive Reputation - Grundzüge einer Reputationstheorie. In U. Röttger, ed. *Theorien der Public Relations: Grundlagen und Perspektiven der PR-Forschung*. 2nd ed. Wiesbaden: Vs Verlag. pp.224-43.

Färber, I. et al., 2010. On Using Class-Labels in Evaluation of Clusterings. In X.Z. Fern, I. Davidson & J. Dy, eds. *MultiClust: Discovering, Summarizing, and Using Multiple Clusterings*. ACM SIGKDD.

Fielding, R.T., 2000. *Architectural Styles and the Design of Network-based Software Architectures*. PhD Thesis. Irvine, CA: University of California, Irvine.

Flyvbjerg, B., 2006. Five Misunderstandings About Case-Study Research. *Qualitative Inquiry, 12 (2)*, April. pp.219-45.

Fombrun, C.J. & van Riel, C.B.M., 2008. *Fame & Fortune: How Successful Companies Build Winning Reputations*. Upper Sadle River NJ: Financial Times Prentice Hall.

Fombrun, C.J. & Wiedmann, K.-P., 2001. Reputation Quotient: Analyse und Gestaltung der Unternehmensreputation auf der Basis fundierter Erkentnisse. *Schriftenreihe Marketing Management*, pp.1-52.

Fuchs, R., 2010. *Kooperieren statt Koordinieren – Web 2.0, Social Software, Wikis: Warum es sich für Unternehmen lohnt, in diesen medientechnologischen Sektor zu investieren*. Licentiat Thesis. Fribourg: University of Fribourg, Departement of Informatics.

Fuhrmann, J. & Wewezow, C., 2010. Herausforderungen und Chancen von Reputationsmanagement im 21. Jahrhundert. In P. Brauckmann, ed. *Web-Monitoring: Gewinnung und Analyse von Daten über das Kommunikationsverhalten im Internet*. 1st ed. Konstanz: UVK Verlagsgesellschaft mbH. pp.363-77.

Fu, L. & Medico, E., 2007. FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data. *BMC Bioinformatics*, 4 January.

Gaines-Ross, L., 2008. *Corporate Reputation: 12 Steps to Safeguarding and Recovering Reputation*. Hoboken: John Wiley & Sons.

Gavirilis, C., Kakali, C. & Papatheodoro, C., 2008. Enhancing Library Services with Web 2.0 Functionalities.

Geisser, S., 1993. *Predictive Inference*. New York: Chapman and Hall.

Glykas, M., ed., 2010. *Fuzzy Cognitive Maps: Advances in Theory, Methodologies, Tools and Applications*. Berlin Heidelberg: Springer Verlag.

Go, A., Bahayani, R. & Huang, L., 2009. *Twitter Sentiment Classification using Distant Supervision*. Stanford: Stanford University.

Goldreich, O., 2008. *Computational Complexity: A Conceptual Perspective*. 1st ed. New York: Cambridge University Press.

Govaert, G., ed., 2009. *Data Analysis*. Hoboken: Wiley.

Griffin, A., 2009. *New Strategies for Reputation Management: Gaining Control of Issues, Crises and Corporate Social Responsibility*. London: Kogan Page.

Gruber, T.R., 1993. A Translation Approach to Portable Ontology Specifications. In *Knowledge Acquisition*., 1993.

Hächler, L., 2010. *Web 2.0 and 3.0: How Online Journalists Find Relevant and Credible Information*. Master Thesis. Fribourg: University of Fribourg, Departement of Informatics.

Hafner, K. & Lyon, M., 2008. *ARPA Kadabra oder Die Anfänge des Internet*. 3rd ed. Heidelberg: dpunkt Verlag.

Halvey, M.J. & Keane, M.T., 2007. An assessment of tag presentation techniques. In *Proceedings of the 16th international conference on World Wide Web*. Banff, Canada, 2007.

Hassan-Montero, Y. & Herrero-Solana, V., 2006. Improving Tag-Clouds as Visual Information Retrieval Interfaces. In *International Conference on Multidisciplinary Information Sciences and Technologies*. Mérida, Spain, 2006.

Hay, D.C., 2002. *Requirements Analysis: From Business Views to Architecture*. Upper Saddle River, New Jersey: Prentice Hall.

Hearst, M., 2011. User Interfaces for Search. In R. Baeza-Yates & B. Ribeiro-Neto, eds. *Modern Information Retrieval: The Concepts and Technology behind search*. 2nd ed. Harlow, England: Pearson Education Ltd. pp.21-55.

Heath, R.L., 1998. New Communication Technologies: An Issues Management Point of View. *Public Relations Review, 24(3)*, pp.273-88.

Heider, F., 1946. Attitudes and Cognitive Organization. *Journal of Psychology 21 (2)*, pp.107-12.

Hein, F.M., 2006. [Online] Available at: http://media1.roadkast.com/interne-kommunikation/Medienstudie_FMHein.pdf [Accessed 4 August 2011].

Hein, F.M., 2009. Bewusster Umgang mit elektronischen Medien in Unternehmen. In A. Back, N. Gronau & K. Tochtermann, eds. *Web 2.0 in der Unternehmenspraxis: Grundlagen, Fallstudien und Trends zum Einsatz von Social Software*. 2nd ed. München: Oldenbourg Wissenschaftsverlag. pp.87-98.

Heinrich, C., Keusen, D., Stuber-Berries, N. & Voutsas, K., 2010. *Der Schweizer Bankenmarkt 2015*. [Online] Accenture Available at: http://www.accenture.com/ch-de/Documents/PDF/Accenture_Banking2015_Studie.pdf [Accessed 15 October 2011].

Herczeg, M., 2009. *Software-Ergonomie: Theorien, Modelle und Kriterien für gebrauchstaugliche interaktive Computersysteme*. 3rd ed. Oldenbourg : Wissenschaftsverlag.

Hevner, A. & Chatterjee, S., 2010. *Design Research in Information Systems: Theory and Practice*. Berlin: Springer.

Hevner, A.R., March, S.T., Park, J. & Ram, S., 2004. Design Science in Information Systems Research. *MIS Quarterly 28(1)*, pp.75-105.

Hexter, E.S. & Bayer, D.S., 2009. Managing Reputation Risk and Reward. *The Conference Board*.

Hitzler, P., Krötzsch, M. & Rudolph, S., 2010. *Foundations of Semantic Web Technologies*. Boca Raton: Chapman & Hall/CRC.

Hitzler, P., Krötzsch, M., Rudolph, S. & Sure, Y., 2008. *Semantic Web: Grundlagen*. Berlin: Springer.

Hohenberg, J., 1978. *The professional Journalist: A Guide to the Practices and Principles of the News Media*. New York: Holt, Rinehart & Winston.

Humphreys, L., 2009. Mobile Social Networks and Services. In T. Dumova & R. Fiordo, eds. *Handbook of Research on Social Interaction Technologies and Collaboration Software: Concepts and Trends*. Hershey: Information Sciences Refernce.

Ibaraki, S. & Lin, A., 2011. *China Value: Social Media for China Business*. [Online] Available at: `http://english.chinavalue.net/AboutUS/TopInterview_Nova_Spivack__World_Renowned__Pioneering_Global_Technology_Visionary__Innovator__Strategist__Entrepreneur__Investor.aspx` [Accessed 26 August 2011].

Ingenhoff, D., 2004. *Corporate Issues Management in multinationalen Unternehmen: Eine empirische Studie zu organisationalen Strukturen und Prozessen*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Ingenhoff, D. & Sommer, K., 2008. The Interrelationships Between Corporate Reputation, Trust and Behavioral Intentions: A Multistakeholder Approach. In *58th Annual Conference of the International Communication Association (ICA)*. Montreal, Canada, 2008.

Jenkins, H., 2008. *Convergence Culture: Where Old and New Media Collide*. Revised ed. New York University Press.

Kaiser, C., 2011. Entscheidungsunterstützung zur Meinungsbeeinflussung in Webcommunitys. In A. Meier & S. Reich, eds. *Communitys im Web*. 48th ed. Heidelberg: HMD. pp.83-93.

Kaser, O. & Lemire, D., 2007. Tag-Cloud Drawing: Algorithms for Cloud Visualization. In *WWW2007 Workshop on Tagging and Metadata for Social Information Organization*. Banff, Alberta, 2007.

Kaufmann, E., 2007. *Talking to the Semantic Web*. Dissertation. Zurich: University of Zurich, Department of Informatics Institut for Informatik.

Kaufmann, M. & Meier, A., 2009. An Inductive Fuzzy Classification Approach applied to Individual Marketing. In *The 28th North American Fuzzy Information Processing Society Annual Conference*. Cincinnati, Ohio, USA, 2009.

Ketchen, D.J. & Shook, C.L., 1996. The application of cluster analysis in Strategic Management Research: An analysis and critique. *Strategic Management Journal* , pp.441-58.

Kitchin, R.M., 1994. Cognitive Maps: What Are They and Why Study Them? *Journal of Environmental Psychology 14 (1)*, p. 1–19.

Klewes, J. & Wreschniok, R., 2009. *Reputation Capital: Building and Maintaining Trust in the 21st Century*. Berlin: Springer.

Klir, G.J., St. Clair, U.H. & Yuan, B., 1997. *Fuzzy Set Theory*. Upper Sadle River: Prentice Hall PTR.

Kolly, S., 2011. *Entwicklung einer unscharfen Ontologie für das Semantische Web*. Bachelor Thesis. Fribourg: University of Fribourg, Departement of Informatics.

Kosko, B., 1986. Fuzzy Cognitive Maps. *International Journal of Man-Machine Studies, 24*, pp. 65-75.

Kruse, R., Döring, C. & Lesot, M.-J., 2007. Fundamentals of Fuzzy Clustering. In J. Valente de Oliveira & W. Pedrycz, eds. *Advances in Fuzzy Clustering and its Applications*. Hoboken: Wiley. pp.3-30.

Kunert, B., Bernet, M. & Allemann, D., 2011. *Bernet PR*. [Online] Bernet PR Available at: `http://www.bernet.ch/images/studies/Social_Media_Studie_Schweiz_Bernet_PR-Kunert.pdf` [Accessed 15 October 2011].

Lämmel, U. & Cleve, J., 2008. *Künstliche Intelligenz*. 3rd ed. München: Carl Hanser Verlag.

Lassila & Swick, 1999. Resource Description Framework (RDF) Model and Syntax Specification, W3C Recommendation, World Wide Web Consortium. Cambridge, MA, 1999.

Levinson, P., 2009. *New new media*. Boston: Alyn and Bacon.

Lewandowski, D., 2005. *Web Information Retrieval: Technologien zur Informationssuche im Internet*. Düsseldorf: Deutsche Gesellschaft f. Informationswissenschaft u. Informationspraxis.

Liechti, A., 2012. *Social Semantic Web: Search Result Ranking for a Fuzzy Reputation Analysis Prototype*. Bachelor Thesis. Fribourg: University of Fribourg, Departement of Informatics.

Lim, Y.-K., Stolterman, E. & Tenenberg, J., 2008. The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. *ACM Transactions on Computer-Human Interaction*, 2 July. pp.7-27.

Lin, N., 2002. *Social Capital: A Theory of Social Structure and Action*. Cambridge: Cambridge University Press.

Lleti, R., Ortiz, M., Sarabia, L. & Sánchez, M., 2004. Selecting Variables for k-Means Cluster Analysis by Using a Genetic Algorithm that Optimises the Silhouettes. *Analytica Chimica Acta*, p.87–100.

Lucko, S. & Trauner, B., 2004. *ABC der Managementtechniken*. München: Carl Hanser Verlag GmbH & CO. KG.

Malone, T.W., 2006. *What is collective intelligence and what will we do about it?* [Online] Available at: `http://cci.mit.edu/about/MaloneLaunchRemarks.html` [Accessed 27 Juni 2011].

Manning, C.D., Raghavan, P. & Schütze, H., 2008. *Introduction to Information Retrieval*. New York: Cambridge University Press.

Manning, C.D. & Schutze, H., 1999. *Foundations of Statistical Natural Language Processing*. MIT Press.

Manovich, L., 2001. *The Language of New Media*. Cambridge: MIT Press.

Mardia, K.V., Kent, T., J. & Bibby, J.M., 1979. *Multivariate Analysis (Probability and mathematical statistics)*. Academic Press.

Marti, S., 2011. *Klassifikation des Reputationsmanagements im Social Semantic Web*. Bachelor Thesis. Fribourg: University of Fribourg, Departement of Informatics.

Martinez, A., 2010. *eGlossar – Ein webbasiertes Glossar unter Nutzung von Web 2.0 und Web 3.0*. Bachelor Thesis. Fribourg, Switzerland: University of Fribourg, Department of Informatics.

Martinez, A., 2011. *eGlossar+: Ein Ontologie-basiertes Glossar*. Seminar Thesis. Fribourg: University of Fribourg, Departement of Informatics.

Maxwell, J.A., 2005. *Qualitative Research Design: An Interactive Approach*. 2nd ed. Thousand Oaks, California: Sage Publications Inc.

McGuinness, D.L. & van Harmelen, F., 2004. *OWL Web Ontology Language Overview*. [Online] Available at: `http://www.w3.org/TR/owl-features/` [Accessed 3 August 2011].

McKay, M., Beckmann, R. & Conover, W., 1979. A comparision of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, pp.239-45.

McLuhan, M. & Nevitt, B., 1972. *Take Today: the Executive as Dropout*. New York.

Meier, A., Schindler, G. & Nicolas, W., 2008. Fuzzy Classification on Relational Databases. In J. Galindo, ed. *Handbook of Research on Fuzzy Information Processing in Databases*. Hershey: IGI Global. pp.586-614.

Meier, A. & Stormer, H., 2009. *eBusiness & eCommerce: Managing the Digital Value Chain*. Berlin: Springer.

Meier, A. & Zumstein, D., 2012. *Web Analytics & Web Controlling: Entwurf, Messung und Nutzung von Webkennzahlen zur Erfolgssicherung*. Heidelberg: dpunkt.verlag GmbH.

Miles, M.B. & Huberman, A.M., 1994. *Qualitative Data Analysis: An Expanded Sourcebook*. 2nd ed. Thousan Oaks: Sage Publications.

Miyamoto, S., Ichihashi, H. & Honda, K., 2008. *Algorithms for Fuzzy Clustering*. Berlin Heidelberg: Springer-Verlag.

Myres, G., 2010. *Discourse of Blogs and Wikis*. London: Continuum International Publishing Group.

Nguyen, H.T. & Walker, E.A., 2005. *A First Course in Fuzzy Logic*. 3rd ed. Boca Raton: Chapman and Hall/CRC.

Nielsen, J., 1989. Usability engineering at a discount. In *Proceedings of the third international conference on human-computer interaction on designing and using human-computer interfaces and knowledge based systems (2nd ed.)*. Boston, Massachusetts, United States, 1989. Elsevier Science Inc.

Novak, J.D., 2010. *Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations*. 2nd ed. New York: Routledge Chapman & Hall.

Novak, J.D. & Cañas, A.J., 2006. *Institute for Human and Machine Cognition*. [Online] Available at: http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMapsHQ.pdf [Accessed 4 August 2011].

Oggier, D., 2009. *Harnessing Folksonomies with a Web Crawler*. Bachelor Thesis. Fribourg: University of Fribourg, Departement of Informatics.

Olivarez-Giles, N., 2011. *Los Angeles Times*. [Online] Available at: http://latimesblogs.latimes.com/technology/2011/09/facebook-f8-media-features.html [Accessed 11 October 2011].

O'Reilly, T., 2005. *What is Web 2.0? Design Patterns and Business Models for the Next Generation of Software*. [Online] Available at: http://oreilly.com/web2/archive/what-is-web-20.html [Accessed 27 Juni 2011].

O'Reilly, T. & Milstein, S., 2009. *The Twitter Book*. Sebastopol: O'Reilly Media.

Orio, N., 2010. Music Indexing and Retrieval for Multimedia Digital Libraries. In M. Agosti, ed. *Information Access through Search Engines and Digital Libraries*. Berlin Heidelberg: Springer. pp.147-69.

Osswald, M., 2011. *Unscharfe, kontextbewusste Ontologien im Social Web*. Bachelor Thesis. Fribourg: University of Fribourg, Departement of Informatics.

Österle, H., Winter, R. & Brenner, W., 2010. *Gestaltungsorientierte Wirtschaftsinformatik: Ein Plädoyer für Rigor und Relevanz*. Nürnberg: Infowerk.

Pedrycz, W. & Gomide, F., 2007. *Fuzzy Systems Engineering: Toward Human-Centric Computing*. 1st ed. Hoboken: John Wiley & Sons.

Pellegrini, T. & Blumauer, A., 2006. Semantic Web und semantische Technologien: Zentrale Begriffe und Unterscheidungen. In T. Pellegrini & A. Blumauer, eds. *Semantic Web: Wege zur vernetzten Wissensgesellschaft*. 1st ed. Berlin Heidelberg: Springer. pp.9-26.

Pepper, S., 2010. *Ontopedia*. [Online] Available at: http://www.ontopedia.net/pepper/papers/ELIS-TopicMaps.pdf [Accessed 4 August 2011].

Peters, P., 2011. *Reputationsmanagement im Social Web*. Norderstedt: Social Media Verlag.

Phillips, L., 2000. *The Double Metaphone Search Algorithm*. [Online] Available at: `http://drdobbs.com/184401251?pgno=2` [Accessed 26 August 2011].

Picot, A., Reichwald, R. & Wigand, R.T., 2003. *Die grenzenlose Unternehmung*. Wiesbaden: Gabler Verlag.

Pilgrim, M., 2010. *HTML5: Up and Running*. 1st ed. Sebastopol: O'Reilly Media;.

Pitts, A., 2009. *Topic Maps Martian Notation*. [Online] Available at: `http://www.musicdna.info/TMMN/` [Accessed 11 October 2011].

Porák, V., Fieseler, C. & Hoffmann, C., 2007. Methoden der Erfolgsmessung von Kommunikation. In M. Piwinger & A. Zerfass, eds. *Handbuch Unternehmenskommunikation*. Wiesbaden: Gabler. pp.535-55.

Portmann, E., 2008. *Informationsextraktion aus Weblogs: Grundlagen und Einsatzmöglichkeiten der gezielten Informationssuche*. Saarbrücken: VDM Verlag.

Portmann, E., 2009. Weblog Extraction with Fuzzy Classification Methods. In *Second International Conference on the Applications of Digital Information and Web Technologies*. London, 2009. IEEE.

Portmann, E., 2011a. A Fuzzy Grassroots Ontology For Improving Social Semantic Web Search. In De Baets, B., Mesiar, R. & Troiano, L., eds. *Proceedings of 6th International Summer School on Aggregation Operators*. Benevento, 2011a.

Portmann, E., 2011b. *Soft Computing - Approximate Reasoning Method for the Semantic Web*. Swiss National Science Foundation (SNSF) application. Fribourg: University of Fribourg, Departement of Informatics.

Portmann, E., Andrushevich, A., Kistler, R. & Klapproth, A., 2010. Prometheus – Fuzzy Information Retrieval for Semantic Homes and Environments. In *International Conference on Human System Interaction*. Rzeszów, 2010.

Portmann, E. & Hutter, R., 2011. Blogosphäre - Soziale Netzwerke für Trendsetter. In A. Meier & S. Reich, eds. *Communitys im Web*. 48th ed. Heidelberg: HMD. pp.37-48.

Portmann, E. & Kuhn, A., 2010. Extraktion und kartographische Visualisierung von Informationen aus Weblogs. In Hengartner, U. & Meier, A. *Web 3.0 & Semantic Web*. Heidelberg: HMD - Praxis der Wirtschaftsinformatik. pp.81-90.

Portmann, E. & Meier, A., 2010. A Fuzzy Grassroots Ontology for improving Weblog Extraction. *Journal of Digital Information Management*, August. pp.276-84.

Portmann, E., Nguyen, T., Sepulveda, J. & Cheok, A.D., 2012. Fuzzy Online Reputation Analysis Framework. In A. Meier & L. Donze, eds. *Fuzzy Methods for Customer Relationship Management and Marketing: Applications and Classification*. Hershey: IGI Global. pp.139 - 167.

Richardson, W., 2010. *Blogs, Wikis, Podcasts, and Other Powerful Web Tools for Classrooms*. 3rd ed. Thousand Oaks, California : Corwin Pr Inc.

Robert, B., 2011. *A Stakeholder's Approach to Issues Management*. Business Expert Press.

Rolke, L. & Köhn, J., 2008. *Medienutzung in der Webgesellschaft 2018: Wie das Internet das Kommunikationsverhalten von Unternehmen, Konsumenten und Medien in Deutschland verändern wird*. Norderstedt: Book on demand Verlag.

Ron, K., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence 2*, p.1137–1143.

Röttger, U., ed., 2001. *Issues Management. Theoretische Konzepte und Praktische Umsetzung. Eine Bestandsaufnahme*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Röttger, U., 2005. Kommunikationsmanagement in der Dualität von Struktur. Die Strukturationstheorie als kommunikationswissenschaftliche Basistheorie. *Medienwissenschaft Schweiz, Nr. 1/2 /2005*, pp.12-19.

Rousseuw, P.J., 1987. Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Computational and Applied Mathematics* , p.53–65.

Saeed, J.I., 2008. *Semantics.* 3rd ed. West Sussex: John Wiley & Sons.

Saenz, A., 2011. *Singularity Hub*. [Online] Available at: `http://singularityhub.com/2011/08/26/mit-unravels-the-`

secrets-behind-collective-intelligence-hint-iq-not-so-important/ [Accessed 27 August 2011].

Sagolla, D., 2009. *140 Characters: A Style Guide for the Short Form*. Hoboken: John Wiley & Sons.

Schreyögg, G. & Koch, J., 2007. *Grundlagen des Managements: Basiswissen für Studium und Praxis*. Wiesbaden: Gabler.

Schwarz, G.E., 1978. Estimating the dimension of a model. *Annals of Statistics*, p.461–464.

Schwede, M. & Stöcklin, D., 2011. *Best Performer Tools im Social-Media-Monitoring: Report 2011*. [Online] Available at: http://www.goldbachinteractive.com/aktuell/fachartikel/social-media-monitoring-report-2011 [Accessed 31 October 2011].

Scott, D.M., 2011. *The New Rules of Marketing & PR: How to Use Social Media, Online Video, Mobile Applications, Blogs, News Releases, and Viral Marketing to Reach Buyers directly*. 3rd ed. Hoboken, New Jersey: John Wiley & Sons.

Seaborne, A. et al., 2008. *SPARQL Update: A language for updating RDF graphs*. [Online] Available at: http://www.w3.org/Submission/SPARQL-Update/ [Accessed 15 October 2011].

Silverman, G., 2011. *The Secrets of Word-of-Mouth Marketing: How to Trigger Exponential Sales Through Runaway Word of Mouth*. 2nd ed. New York: Mcgraw-Hill Professional.

Simon, H.A., 1996. *The Sciences of the Artificial*. 3rd ed. Cambridge, MA: MIT Press.

Simou, N. & Kollias, S., 2007. FiRE: A Fuzzy Reasoning Engine for Imprecise Knowledge. Berlin, 2007.

Sirmakessis, S., 2005. *Knowledge Mining: Proceedings of the NEMIS 2004 Final Conference*. Berlin Heidelberg: Springer.

Smith, G., 2008. *Tagging: People-Powered Metadata for the Social Web*. Berkeley: New Riders.

Smithson, M. & Verkuilen, J., 2006. *Fuzzy Set Theory*. Thousand Oaks: Sage Publications, Inc.

Snyder, C., 2003. *Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces*. San Francisco: Morgan Kaufmann.

Sooth, S. & Schoneville, C., 2011. *WIKIPEDIA for World Heritage*. [Online] Available at: `http://wikipedia.de/wke/Main_Page` [Accessed 26 August 2011].

Spitzer, M., 2000. *Geist im Netz: Modelle für Lernen, Denken und Handeln*. Heidelberg: Spektrum Verlag.

Stegbauer, C., 2009. *Wikipedia: Das Rätsel der Kooperation*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Sterne, J., 2011. *Social Media Metrics: How to measure and optimize your Marketing Investment*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Stoilos, G., Simou, N., Stamou, G. & Kollias, S., 2006. Uncertainty and the Semantic Web. *IEEE Intelligent Systems*, 2 October. pp.84-87.

Stoilos, G. et al., 2005a. The fuzzy description logic f-SHIN. In *Proc. of the International Workshop on Uncertainty Reasoning for the Semantic Web*., 2005a.

Stoilos, G. et al., 2005b. Fuzzy OWL: Uncertainty and the Semantic Web. In *Proc. of the International Workshop on OWL: Experiences and Directions*., 2005b.

Stoilos, G. et al., 2005. Fuzzy OWLl: Uncertainty and the Semantic Web. In *Proc. of the International Workshop on OWL: Experiences and Directions*., 2005.

Sugar, C.A. & James, G.M., 2003. Finding the number of clusters in a data set: An information theoretic approach. *Journal of the American Statistical Association*, January. p.750–763.

Suzuki, E. & Setsuo, A., 2004. Discorvery Science. In *7th International Conference, DS 2004*. Padova, Italy, 2004.

Thiessen, A., 2011. *Organisationskommunikation in Krisen: Reputationsmanagement durch situative, integrierte und strategische Krisenkommunikation*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Thomas, G., 2010. *How to do your Case Study: A Guide for Students and Researchers*. London: Sage Publications Ltd.

Tiwari, S., 2011. *Professional NoSQL*. 1st ed. John Wiley & Sons.

Toffler, A., 1980. *The Third Wave*. Collins.

Too, Y.L., 2010. *The Idea of the Library in the Ancient World*. New York: Oxford University Press.

Uhlmann, K., 2011. *Klassifikation und Einordnung der Online Reputationsanalyse in Schweizer Finanzdienstleistungsinstituten*. Master Thesis. Fribourg: University of Fribourg, Departement of Informatics.

Vaishnavi, V. & Kuechler, B., 2009. *Design Research in Information Systems*. [Online] Available at: `http://desrist.org/design-research-in-information-systems/` [Accessed 26 August 2011].

van Aken, J.E., 2004. Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules. *Journal of Management Studies*, March. pp.219-46.

van Gaalen, M., 2009. *ORM: more than just setting things right*. [Online] Available at: `http://www.jungleminds.com/publications/articles/orm%3A_more_than_just_setting_things_right/` [Accessed 4 August 2011].

van Harmelen, F., Lifschitz, V. & Porter, B., 2007. *Handbook of Knowledge Representation*. New York: Elsevier.

van Riel, C. & Fombrun, C.J., 2007. *Essentials of Corporate Communication: Implementing practices for effective reputation management*. Abingdon: Routledge.

Venkata Rao, R., 2010. *Decision Making in the Manufacturing Environment: Using Graph Theory and Fuzzy Multiple Attribute Decision Making Methods*. Berlin: Springer.

Voss, J., 2007. Tagging, Folksonomy & Co – Renaissance of Manual Indexing? In *10th International Symposium for Information Science*. Cologne, 2007.

Walsh, G., 2006. *Das Management von Unternehmensreputation: Grundlagen, Messung und Gestaltungsperspektiven am Beispiel von Unternehmen des liberalisierten Gasmarkts*. 1st ed. Aachen: Shaker.

Ward, M., Grinstein, G.G. & Keim, D., 2010. *Interactive Data Visualization: Foundations, Techniques, and Applications*. Natick, MA: Taylor & Francis Ltd.

Ward, M., Grinstein, G.G. & Keim, D., 2010. *Interactive Data Visualization: Foundations, Techniques, and Applications*. Natick: Taylor & Francis Ltd.

Wehrle, M. & Portmann, E., 2012. Monitor for Nanotechnology risks. *Second International Conference on Advanced Communications and Computation*.

Weick, K.E. & Sutcliffe, K.M., 2007. *Managing the Unexpected: Resilient Performance in an Age of Uncertainty*. 2nd ed. San Francisco: John Wiley & Sons.

Weller, K., 2010. *Knowledge Representation in the Social Semantic Web*. Berlin/New York: de Gruyter Saur Verlag.

Wilde, T. & Hess, T., 2007. Forschungsmethoden der Wirtschaftsinformatik. In *Wirtschaftsinformatik*. Gabler Verlag. pp.280-87.

Wilson, R., 2005. Reputations in Games and Markets. In A.E. Roth, ed. *Game Theoretic Models of Bargaining*. Cambridge: Cambridge University Press.

Zacklad, M., Cahier, J.-P. & Pétard, X., 2003. *Du Web Cognitivement Sémantique au Web Socio Sémantique*. Journée. Paris: Web Sémantique et Sciences humaines et sociales.

Zadeh, L.A., 1965. Fuzzy Sets. *Information and Control*, 8(3), p.338–353.

Zadeh, L.A., 1994. Fuzzy Logic, Neural Networks, and Soft Computing. *Communications of the ACM*, March. pp.77-84.

Zadeh, L.A., 1998. Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems. In *Soft Computing*. 2nd ed. Springer-Verlag. pp.23-25.

Zadeh, L.A., 2002. From Computing with Numbers to Computing with Words—From Manipulation of Measurements to Manipulation of Perceptions. *International Journal of Applied Math and Computer Science, Vol.12, No.3,* , p.307–324.

Zadeh, L.A., 2004. A note on web intelligence, world knowledge and fuzzy logic. *Data & Knowledge Engineering*, September. pp.291-304.

Zadeh, L.A., 2006. From Search Engines to Question Answering Systems – The Problems of World Knowledge, Relevance, Deduction and Precisiation. In E. Sanchez, ed. *Fuzzy Logic and the Semantic Web*. Amsterdam: Elsevier Ltd. pp.163-210.

Zadeh, L.A., 2009. Toward extended fuzzy logic—A first step. In *Fuzzy Sets and Systems*. pp.3175-81.

Zadeh, L., 2010. *BISC: The Berkeley Initiative in Soft Computing*. [Online] UC Berkeley Available at: `http://www.cs.berkeley.edu/~zadeh/presentations%202008/SSSC%2710-Full%20version-Precisiation%20of%20Meaning-July26.pdf` [Accessed 26 August 2011].

Zerfass, A., 2004. *Unternehmensführung und Öffentlichkeitsarbeit. Grundlegung einer Theorie der Unternehmenskommunikation und Public Relations*. 2nd ed. Wiesbaden: VS Verlag für Sozialwissenschaften.

Zimmermann, H.-J., 2001. *Fuzzy Set Theory - and its Applications*. 4th ed. Boston/Dordrecht/London: Kluwer Academic Publisher.

Zudilova-Seinstra, E., Adriaansen, T. & Van Liere, R., 2008. *Trends in Interactive Visualization: State-of-the-Art Survey*. Berlin: Springer.

GLOSSARY

+1 Button: Is a feature to recommend Web content.

ABox: Describes in conjunctive with →TBox two different statements in →Ontologies. Together they make up a →Knowledge Base.

ACID: The principles of Atomicity, Consistency, Isolation, and Durability guarantee that database transactions are processes reliably.

Actor Network Theory: See →ANT.

Another Neat Tool: See →ANT.

ANT: Has multiple meanings. Here either Actor Network Theory that is an approach to social theory and research or Apaches Another Neat Tool for automating software build processes.

API: An Application Programming Interface is a specification that software can act on to communicate with each other.

APP: The Atom Publishing Protocol is a specification for creating and updating Web content.

Applet: Is a small Web application that is used for interactivity that cannot be provided by →HTML.

Application Programming Interface: See →API.

Approximate Reasoning Methods: See →ARM.

ARM: The Approximate Reasoning Methods project is a →SNSF-granted fellowship with the aim to better integrate →Fuzziness in →Social Semantic Web.

Artificial Intelligence: Is the intelligence of computers. It includes →Knowledge Representation and Reasoning, →Machine Learning, and →NLP.

Atomicity, Consistency, Isolation, and Durability: See →ACID.

Atom Publishing Protocol: See →APP.

Azure: A →Cloud Computing service.

Basic Logic Dialect: See →BLD.

BLD: The Basic Logic Dialect adds features to the →RIF →Core Dialect that are not directly available.

Blog: See →Weblog and →Microblog.

Blogger: A person who writes a →Blog, or a Web service for publishing →Blogs.

Bookmark: A stored →URI with the possibility to annex →Metadata.

Cascading Style Sheets: See →CSS.

CCO: The Chief Communications Officer is the highest-ranking executive in charge of an organization's communications and →PR.

CEO: The Chief Executive Officer is the highest-ranking executive in charge of an organization's total management.

CFO: The Chief Financial Officer is the highest-ranking executive in charge of an organization's financial risk management.

Chief Communication Officer: See →CCO.

Chief Executive Officer: See →CEO.

Chief Financial Officer: See →CFO.

Chief Web Officer: See →CWO.

Cloud Computing: Is computing as a service rather than a product, whereby resources, software, and →Information are provided over a network (typically the →Internet). It includes →Azure, →iCloud, and →Ubuntu One.

Cluster Analysis: See →Clustering.

Cluster Supporting Objects: →See CSO.

Cluster: Is a bunch of elements in a set.

Clustering: Is the task of assigning elements into →Clusters so that the elements in the same →Cluster are more similar to each other than to those in other →Clusters.

CMS: A Content Management System provides a collection of procedures used to manage workflow in a collaborative environment.

Collective Intelligence: Is a group aptitude that emerges from the collaboration of many different individuals.

Communication Operatives: Employees concerned with an organizations' communication. They are either subject to the →CCO or the →CWO.

Content Management System: See →CMS.

Context: The relevant constraints of a situation.

Core Dialect: Is a subset of most rule dialects as →RIF →BLD and →RIF →PRD.

CRM: Customer Relationship Management is a strategy for managing an organization's interactions with its stakeholders.

Crowd: Is a group of individuals remaining united through a common purpose.

Crowdsourcing: A blend of →Crowd and outsourcing is the contracting out of tasks to a group of individuals.

CSO: Cluster Supporting Objects are archetypal elements around which a →Cluster will be constructed.

CSS: Cascading Style Sheets are used to describe the formatting of →Web →Data written in a →Markup Language.

Customer Relationship Management: See →CRM.

CWO: The Chief Web Officer is the highest-ranking executive in

charge of an organizations Web presence.

Data: The lowest level of abstraction from which →Information and →Knowledge can be derived.

Delicious: Is a →Folksonomy service for storing, sharing, and discovering →Social Bookmarks.

Description Logic: See →DL.

Description Of A Project: See →DOAP.

DL: Description Logic is a formal →Knowledge Representation language for formal reasoning.

DOAP: Description Of A Project is a vocabulary to describe software projects.

eGlossary: The →Electronic Glossary is a →Social Semantic Web dictionary that presents explanations syntonic to individuals state of →Knowledge.

Electronic Glossary: See →eGlossary.

Enhanced Collective Intelligence: Integrates computers' aptitude in →Collective Intelligence.

eXtensible Markup Language: See →XML.

Facebook: Is a →Social Network service for interacting with online friends.

FCM: The Fuzzy C-Means algorithm is a →Fuzzy Clustering algorithm to improve the accuracy of →Clustering under →Fuzziness.

FLAME: The Fuzzy clustering by Local Approximation of MEmbership algorithm is a →Fuzzy Clustering algorithm that allows capturing non-linear relationships and non-globular →Clusters, an automated definition of →Cluster numbers and the identification of outliers in a set.

Flickr: Is an image and video hosting service to share and embed personal Web content to →Social Media Elements.

FOAF: Friend-Of-A-Friend is an →Ontology describing individuals, their activities and their relations to other individuals and objects.

Folksonomy: A blend of folk and taxonomy is a →Social Media Element. It comprises a →Social Bookmarking system that depends on →Collective Intelligence to →Tag Web content.

FORA Framework: The Fuzzy Online Reputation Analysis Framework is a →Online Reputation Analysis frame that helps handle →Fuzziness in →Online Reputation Management process.

Forecasting: Based on history →Data this technique attempts to predict future trends.

Friend-Of-A-Friend: See →FOAF.

Fuzziness: Vagueness that arises in real-world complexity.

Fuzzy clustering by Local Approximation of MEmberships: See →FLAME.

Fuzzy Clustering: In this →Clustering method to handle →Fuzziness each element has a →Membership Degree to →Clusters rather than belonging completely to just one →Cluster.

Fuzzy C-Means: See →FCM.

Fuzzy Grassroots Ontology: An →Ontology spawned from →Folksonomy to react to →Fuzziness.

Fuzzy Logic: Deals with reasoning that is approximate rather than fixed and exact to handle →Fuzziness.

Fuzzy Online Reputation Analysis Framework: See →FORA Framework.

Fuzzy Set Theory: Sets whose elements have different →Membership Degrees of belonging to different sets to handle →Fuzziness.

Fuzzy: Helps handle →Fuzziness and refers to →Fuzzy Logic, →Fuzzy Set Theory, and →Fuzzy Clustering.

GK: The Gustafson-Kessel is a →Fuzzy Clustering algorithm that allows detecting →Clusters of different geometrical shapes in a set.

Google Alerts: Is a Web content change detection and notification service.

Google Analytics: Is a service that generates statistics about the visitors of a website.

Google Find my Face: Is a service that recognizes and annotate friends on online photos.

Google Maps: Is a service that offers map-based services.

Google+: Is a →Social Network service for interacting with online friends.

Graph Database: A database that uses graph structures with nodes, edges, and properties to represent and store →Knowledge. It includes →Triple Stores.

Graphical User Interface: See →GUI.

GUI: A Graphical User Interface allows interacting with computers and mobile devices with images rather than text commands.

Gustafson-Kessel: See →GK.

Hashtag: Are means to →Tag →Web content and may provide a →Folksonomy.

HCI: Human-Computer Interaction is the study, planning and design of the interaction between humans and computers.

HCIR: Human-Computer Information Retrieval combines →IR and →HCI, in order to create search engines that depend on continuous human control of the search process.

HITS: The Hyperlink-Induced Topic Search algorithm rates links in Web content.

Homograph: Words that share the same spelling, regardless of how they are pronounced.

Homonym: Words that share the same spelling and the same pronunciation but have different meanings.

Homophone: Words that share the same pronunciation, regardless of how they are spelled.

HTML: The Hypertext Markup Language is a →Markup Language for Web content.

HTTP: The Hypertext Transfer Protocol is the foundation of Web communication.

Human-Computer Information Retrieval: See →HCIR.

Human-Computer Interaction: See →HCI.

Hyperlinked-Induced Topic Search: See →HITS.

HyperText Markup Language: See →HTML.

HyperText Transfer Protocol: See →HTTP.

iCloud: Is a →Cloud Computing service.

Information Overload: Refers to the difficulty of understanding and making decisions caused by too much →Information.

Information Retrieval: See →IR.

Information: Evaluated →Data from which →Knowledge can be derived.

Innovation- and Risk-Navigator: See →InRiNa.

InRiNa: The Innovation- and Risk-Navigator project encourages nanotechnology specialists to identify more relevant →Information during a search using a →Fuzzy Grassroots Ontology.

Interactive Visualization: Reciprocal graphic representation of →Information that involves →HCI.

International Standardization Organization: See →ISO.

Internet: A global system of interconnected networks. It includes the →WWW.

Intranet: A network specification to securely share →Information within an organization.

iPad: Is a tablet computer for managing multimedia content. Its size and weight falls between those of smartphones and laptop computers.

iPhone: Is a multimedia-enabled smartphone.

iPod: Is a portable media player.

IR: Information Retrieval is concerned with searching for →Data, for →Metadata about →Data, and for →Information within →Data.

ISO: The International Standardization Organization is an international standards organization composed of representatives from national standards organizations.

JavaScript Object Notation: See →JSON.

JavaScript: Is a dynamic scripting language.

JSON: →JavaScript Object Notation, is a standard for representing →Data structures.

KAON: The KArlsruhe ONtology is an →Ontology infrastructure developed by the University of Karlsruhe (Germany).

KArlsruhe ONtology: See →KAON.

KNN: The k-Nearest Neighbors are the nearest elements with high similarity.

Knowledge Administration System: Based on a →Knowledge Base it provides the means for the computerized →Knowledge Discovery. It includes →Graph Databases.

Knowledge Base: A collection of instances of concepts defined in a →Ontology. It is also a special kind of →Knowledge Administration System.

Knowledge Discovery: A concept that describes the process of automatically collecting, organizing, and retrieving →Knowledge.

Knowledge Representation (and Reasoning): Represents →Knowledge in symbols to facilitate inferencing from those, creating new →Knowledge.

Knowledge: A familiarity with something that involves →Information, acquired through experience or education.

k-Nearest Neighbor: See →KNN.

Last.fm: Is a →Social Networking and music recommender service.

Like Button: Is a feature of →Facebook to recommend Web content.

LinkedIn: Is a business-related →Social Network service.

Machine Learning: Is concerned with the design and development of software that allows computers to evolve behaviors based on →Data.

Markup Language: A system for annotating a text in a way that is syntactically distinguishable from that text.

Mediamatics: A blend of multimedia and informatics that uses a combination of different content forms to ease →HCI.

Membership Degree: Is the gradual assessment of membership of elements in a set.

Metadata: Is →Data providing further →Information about one or more aspects of specific →Data.

Metaphone: Is a phonetic algorithm for indexing words by their pronunciation.

Microblog: A blend of micro and →Weblog is a →Social Media Element. It is a →Weblog with smaller content in both actual and aggregate file size.

Monitoring: Is the continuous and targeted observation of identified issues.

National University of Singapore: See →NUS.

Natural Language Processing: See →NLP.

NLP: Natural Language Processing is concerned with →HCI in human natural languages.

NUS: The National University of Singapore is Singapore's oldest University.

Online Reputation Analysis: A →Online Reputation Management task conducted by communications operatives. Consists of the practice of →Scanning, →Monitoring, and →Forecasting.

Online Reputation Management: Is the act of addressing or rectifying undesirable or negative →SERPs or mentions in →Social Media.

Ontology: Stands for a design model for specifying the world that consists of a set of types, relationships and properties.

Open-World Assumption: See →OWA.

OWA: The Open World Assumption is the hypothesis that the truth of a statement is independent of whether or not it is known to be true. This stands in contrast to the hypothesis, which holds, that any statement that is not known to be true is false.

OWL: The Web Ontology Language is a →W3C standardized →Knowledge Representation languages for describing and sharing →Ontologies on the →WWW.

Paradox: A seemingly true statement that leads to a contradiction or a situation that seems to defy logic or intuition.

PDA: A Personal Digital Assistant is a mobile device that functions as an individual →Information manager.

Permalink: A blend of permanent and link is a →URL that points to a specific →Blog post.

Personal Digital Assistant: See →PDA.

PHP Hypertext Preprocessor: See →PHP.

PHP: The PHP Hypertext Preprocessor is free software to produce dynamic webpages.

PR: Public Relations is the practice of managing communication between an organization and its stakeholders.

PRD: The Production Rules Dialect adds features to the →RIF →Core Dialect that are not directly available.

Pre-Media Space: Encompasses the area between individuals, organizations and (traditional) media.

Production Rule Dialect: See →PRD.

Prometheus Framework: A building intelligence project that is based on a →Fuzzy Grassroots Ontology. It encourages its users during a context-aware search.

Prosumer: A blend of producer and consumer that illustrates the

personae of individual in the →WWW.

Proximity: Numerical descriptions of how far apart elements are.

Public Relations: See →PR.

RAM: Random Access Memory is a form of computer data storage.

Random-Access Memory: See →RAM.

RDF in Attributes: See →RDFa.

RDF Schema: See →RDFS.

RDF: The Resource Description Framework is a →W3C standard designed as a →Metadata model.

RDFa: RDF in Attributes is a →W3C recommendation for embedding →Metadata within Web content.

RDFS: The RDF Schema is a →W3C standardized →Knowledge Representation language providing basic elements for describing →Ontologies.

Really Simple Syndication: See →RSS.

REpresentational State Transfer: See →REST.

Resource Description Framework: See →RDF.

REST: The REpresentational State Transfer is software architecture for distributed hypermedia systems such as the →WWW.

RIF: The Rule Interchange Format is a →W3C recommendation that allows exchanging rules between different rule languages. It includes the →Core Dialect, as well as →BLD, and →PRD.

RSS: The RDF Site Summary is a family of Web Feed formats used to publish frequently updated Web content in a standardized format.

Rule Interchange Format: See →RIF.

Scanning: Is an early detection of changes in the environment of an organization that may restrict its scope.

Search Engine Optimization: See →SEO.

Search Engine Result Pages: See →SERP.

Semantic Web: A Web of →Data that facilitates computers to understand the →Semantics of →Information on the →WWW.

Semantically-Interlinked Online Communities: See →SIOC.

Sentiment Analysis: Use →NLP to identify and extract subjective →Information in Web content.

SEO: Search Engine Optimization is the practice of improving the visibility of Web content or a webpage in →SERPs.

SERP: A Search Engine Results Page is the listing of webpages returned by a search engine answering a keyword query.

Short Messages Service: See →SMS.

SIOC: The Semantically-Interlinked Online Communities project is a method for interconnecting discussions to each other.

Small and Medium Enterprises: See →SME.

SME: Small and Medium Enterprises are organizations whose headcount or turnover falls below certain criteria.

SMS: Short Message Service is a messaging service using standardized communications protocols that allow the exchange of short text messages.

SNSF: The Swiss National Science Foundation is a Swiss board promoting scientific research.

Social Bookmark: A shared →Bookmark possibly resulting in a →Folksonomy.

Social Media Elements: The diverse manifestations of →Social Media as technical implementations. It includes →Folksonomies, →Microblogs, →Social Networks, →Weblogs, and →Wikis.

Social Media: Refers to the use of →Web and mobile devices to turn communication into an interactive dialogue.

Social Network: A form of →Social Media Element. It comprises an online platform that focuses on building and reflecting of social relations among individuals.

Social Semantic Web: Subsumes developments in which social interactions on the →WWW lead to the creation of semantically rich →Knowledge Representations. It combines technologies, strategies and methodologies from the →Social Web with the →Semantic Web.

Social Web: People socializing with each through →Social Media Elements.

Soft Computing: Use of inexact solutions to computationally-hard tasks. Includes →Fuzzy Logic.

SPARQL Protocol And RDF Query Language: See →SPARQL.

SPARQL: The SPARQL Protocol and →RDF Query Language is a →W3C standardized →RDF query language.

SQL: The Structured Query Language is a standardized relational database query language.

Structured Query Language: See →SQL.

Swiss National Science Foundation: See →SNSF.

Synonym: Words with almost identical or similar meanings.

Tag Cloud: A visual representation for text data, used to depict tags on Web content, or to visualize free form text.

Tag: Non-hierarchical →Metadata assigned to →Information.

Tagspace: Is a set of associated →Tags with related weights.

TBox: Describes in conjunction with →ABox two different state-

ments in →Ontologies. Together they make up a →Knowledge Base.

TMMN: The Topic Map Martian Notation is a graphical notation used to explain the →Topic Maps data model, and map out →Ontologies and instance data.

Topic Map Martian Notation: See →TMMN.

Topic Map: A →ISO standard of →Knowledge Representation language with an emphasis on the findability of information.

Triple Store: A database for the storage and retrieval of →RDF.

Tumblr: Is a Web service for publishing →Blogs.

Twitter: Is a →Microblogging and →Social Networking service that enables users to send and read short real-time messages.

Ubuntu One: Is a →Cloud Computing service that enables users to store and sync files online and between computers.

UGC: User Generated Content is individual-created →Social Media about which the providers of a certain webpage do not care.

UML: The Unified Modeling Language is a standardized object specification language used in software engineering.

Unicode: A standard for the consistent encoding, representation and handling of text.

Unified Modeling Language: See →UML.

Uniform Resource Identifier: See →URI.

Uniform Resource Locator: See →URL.

URI: A Uniform Resource Identifier is a character string used to identify content on the →Internet.

URL: A Uniform Resource Locator is a type of an URI that helps to retrieve content on the →Internet.

User Generated Content: See →UGC.

W3C: The World Wide Web Consortium is an international standards organization for the →WWW.

Web Agent: An application used in communications within a client–server distributed computing system. It browses the →WWW in a methodical and automated manner.

Web Browser: An application used for retrieving, presenting, and traversing Web content on the →WWW.

Web Feed: A format used for providing frequently updated Web content. It includes →APP and →RSS.

Web of Things: See →WOT.

Web Ontology Language: See →OWL.

Web Services Description Language: See →WSDL.

Web: See →WWW.

Weblog: A blend of Web and log is a →Social Media Element. It compromises a regularly updated website. The ability to leave comments in an interactive format constitutes an important part of it.

WEF: The World Economic Forum is a Swiss board holding annual meetings of world economic leaders in the eastern Alps region of Switzerland.

Wiki: A →Social Media Element that allows creating and editing of webpages via a →Web Browser using a simplified →Markup Language. It includes →Wikipedia, →Wikiquote, and →Wikitravel.

Wikipedia: Is a →Wiki service that goes as a free, Web-based, collaborative, multilingual encyclopedia.

Wikiquote: Is a →Wiki service that provides reference of quotations.

Wikitravel: Is a →Wiki service that goes as a Web-based collaborative travel guide.

WOM: Word Of Mouth is the credible viral passing of →Information from individual to individual and counts in the →WWW as a personal recommendation.

Word Of Mouth: See →WOM.

WordPress: A Web service for publishing →Blogs.

World Economic Forum: See →WEF.

World Wide Web Consortium: See →W3C.

World Wide Web: See →WWW.

WOT: The Web of Things is a vision where computers, mobile devices and everyday objects are integrated into the →WWW using Web standards.

WSDL: The Web Services Description Language is a →W3C recommendation that constitutes a →XML-based language used for describing a Web service.

WWW: The World Wide Web, or Web for short, is a system of interlinked hypertext documents accessed via the →Internet.

XML Topic Maps: See →XTM.

XML: The eXtensible Markup Language is a →W3C standard for encoding documents in computer-readable form.

XTM: XML Topic Maps are →Topic Maps for the →WWW represented with the aid of →XML.

YouReputation Prototype: See →YouReputation.

YouReputation: A blend of your and reputation is a free and easy-to-use →Online Reputation Analysis Tool. It goes as proof of concept for the →FORA Framework.

CURRICULUM VITAE

PERSONAL INFORMATION

- *Full Name*: Eduard Karl PORTMANN
- *Birthday*: August 13, 1976
- *Citizenship*: Swiss (from Ruswil LU)
- *Marital status*: married, one daughter

EDUCATION

- *2009–2012*: Ph.D. in Computer Science at University of Fribourg (Switzerland)
- *2007–2009*: M.Sc. in Business & Economics at University of Basel[45] (Switzerland)
- *1999-2003*: B.Sc. in Information Systems at Lucerne University of Applied Sciences and Arts[46] (Switzerland)

WORK EXPERIENCE

- *Summer 2010*: Visiting Research Scholar at National University of Singapore (Singapore)
- *2008–2010*: Researcher at Lucerne University of Applied Sciences and Arts (Switzerland)
- *2006*: IT-Auditor at Ernst & Young[47], Technology and Security Risk Services (Switzerland)
- *2005–2006*: Business Analyst at PricewaterhouseCoopers[48], Global Technology Solutions (Switzerland)
- *2004–2005*: Contract Manager at Swisscom Mobile[49], Business Steering (Switzerland)
- *2003–2004*: Supervisor at Link Market Research Institute[50] (Switzerland)

PUBLICATIONS

- Wehrle, M. & Portmann, E., 2012. Monitor for Nanotechnology risks. Second International Conference on Advanced Communications and Computation.

- Portmann, E., Nguyen, T., Sepulveda, J. & Cheok, A.D., 2012. Fuzzy Online Reputation Analysis Framework. In A. Meier & L. Donze, eds. Fuzzy Methods for Customer Relationship Management and Marketing: Applications and Classification. Hershey: IGI Global. pp.139-167.

- Portmann, E. & Hutter, R., 2011. Blogosphäre - Soziale Netzwerke für Trendsetter. In A. Meier & S. Reich, eds. Communitys im Web. 48th ed. Heidelberg: HMD. pp.37-48.

- Andrushevich, A. et al., 2011. Prometheus Framework for Fuzzy Information Retrieval in Semantic Spaces. In Human Computer Systems Interaction: Backgrounds and Applications. 2nd ed. Springer.

- Drobnjak, A. et al., 2011. Führungsinformationssysteme unter Nutzung der unscharfen Logik - Fallbeispiel coop@home, internal working paper, 2011.

- Portmann, E., 2011a. A Fuzzy Grassroots Ontology For Improving Social Semantic Web Search. In De Baets, B., Mesiar, R. & Troiano, L., eds. Proceedings of 6th International Summer School on Aggregation Operators. Benevento, 2011a.

- Portmann, E., 2011b. Soft Computing-Approximate Reasoning Method for the Semantic Web. Swiss National Science Foundation (SNSF) application. Fribourg: University of Fribourg.

- Portmann, E. & Meier, A., 2010. A Fuzzy Grassroots Ontology for improving Weblog Extraction. Journal of Digital Information Management, August. pp.276-84.

- Portmann, E. & Kuhn, A., 2010. Extraktion und kartographische Visualisierung von Informationen aus Weblogs. In Hengartner, U. & Meier, A. Web 3.0 & Semantic Web. Heidelberg: HMD - Praxis der Wirtschaftsinformatik. pp.81-90.

- Portmann, E., Andrushevich, A., Kistler, R. & Klapproth, A., 2010. Prometheus – Fuzzy Information Retrieval for Semantic Homes and Environments. In International Conference on Human System Interaction. Rzeszów, 2010.

- Portmann, E., 2009. Weblog Extraction with Fuzzy Classification Methods. In Second International Conference on the Applications of Digital Information and Web Technologies. London, 2009. IEEE.

- Portmann, E., 2008. Informationsextraktion aus Weblogs: Grundlagen und Einsatzmöglichkeiten der gezielten Informationssuche. Saarbrücken: VDM Verlag.

ENDNOTES

The respective website can be reached at:

[1] http://www.blogger.com/
[2] http://twitter.com/
[3] http://wordpress.com/
[4] http://www.flickr.com/
[5] http://www.last.fm/
[6] http://delicious.com/
[7] http://www.facebook.com/
[8] http://plus.google.com/
[9] http://www.linkedin.com/
[10] http://www.wikipedia.org/
[11] http://wikitravel.org/
[12] http://www.wikiquote.org/
[13] http://www.youreputation.org/
[14] http://www.apple.com/
[15] http://www.weforum.org/
[16] http://www.tumblr.com/
[17] http://www.google.com/
[18] http://www.apple.com/icloud/
[19] http://dublincore.org/
[20] http://dbpedia.org/
[21] http://semantic-mediawiki.org/
[22] http://www.foaf-project.org/
[23] http://trac.usefulinc.com/doap
[24] http://sioc-project.org/
[25] http://inrina.com/
[26] http://www.ihomelab.ch/
[27] http://www.google.com/analytics/
[28] http://www.microsoft.com/
[29] http://www.google.com/alerts
[30] http://www.unifr.ch/
[31] http://www.microsoft.com/windowsazure/
[32] http://one.ubuntu.com/
[33] http://www.kit.edu/
[34] http://www.google.com/maps
[35] http://diuf.unifr.ch/is/
[36] http://www.cutecenter.org/
[37] http://www.nus.edu.sg/
[38] http://www.keio.ac.jp/
[39] http://drupal.org/
[40] http://wordnet.princeton.edu/

[41] http://www.sysomos.com/
[42] http://www.memonews.com/
[43] http://www.actionly.com/
[44] http://www.snf.ch/
[45] http://www.unibas.ch/
[46] http://www.hslu.ch/
[47] http://www.ey.com/
[48] http://www.pwc.com/
[49] http://www.swisscom.ch/
[50] http://www.link.ch/