Annals of Information Systems

Vladan Devedžić • Dragan Gašević
Editors

Web 2.0 & Semantic Web

Springer

Editors
Vladan Devedžić
University of Belgrade
School of Business Administration
Dept. Information Systems & Technologies
Jove Ilica 154
11000 Belgrade
Serbia
devedzic@fon.rs

Dragan Gašević
Athabasca University
School of Computing
& Information Systems
1 University Drive
Athabasca AB T9S 3A3
Canada
dragang@athabascau.ca

ISSN 1934-3221 e-ISSN 1934-3213
ISBN 978-1-4419-1218-3 e-ISBN 978-1-4419-1219-0
DOI 10.1007/978-1-4419-1219-0
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2009933257

Preface

According to the W3C Semantic Web Activity [1]: *The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries.* This statement clearly explains that the Semantic Web is about data sharing. Currently, the Web uses hyperlinks to connect Web pages. The Semantic Web goes beyond that and focuses on *data* and envisions the creation of the *web of data*. On the Semantic Web, anyone can say anything about any resource on the Web. This is fully based on the concept of semantic annotations, where each resource on the Web can have an assigned meaning. This is done through the use of ontologies as a formal and explicit representation of domain concepts and their relationships [2]. Ontologies are formally based on description logics. This enables agents and applications to reason over the data when searching the Web, which has not previously been possible.

Web 2.0 has gradually evolved from letting the Web users play a more active role. Unlike the initial version of the Web, where the users mainly "consumed" content, users are now offered easy-to-use services for content production and publication. Mashups, blogs, wikis, feeds, interface remixes, and social networking/tagging systems are examples of these well-known services. The success and wide adoption of Web 2.0 was in its reliance on social interactions as an inevitable characteristic of the use and life of the Web. In particular, Web 2.0 focuses on creating knowledge through collaboration and the social interactions of individuals (e.g., wikis). These systems use terms (*tags*) to reflect personal assertions about resources, recommend content to the other members in the community, as well as to build a shared community vocabulary (*folksonomy*).

Both Web 2.0 and the Semantic Web obviously offer many benefits, but at the same time exhibit some deficiencies. On the one hand, the Semantic Web requires very expensive knowledge acquisition procedures in order to make use of its full power. Examples are expert involvement in ontology development and advanced semantic annotation techniques. The recent research on ontologies suggests that ontologies are not just about symbols representing knowledge, but also about the social interactions of the ontology users [3]. This notion has considerable influence on the adoption of Semantic Web technologies, as the construction, use, and evolution of

ontologies and semantic annotation are difficult tasks [4–6]. On the other hand, Web 2.0 technologies in general, and collaborative tagging in particular, suffer from the problems of ambiguity in their tags' meanings and the lack of semantics (e.g., synonyms), the lack of coherent categorization schemata, and the needed time and size of the community in which they will be used [7]. Intuitively, this can be addressed by ontologies, clearly explaining why the Semantic Web and Web 2.0 are complementary approaches often referred to as the Social Semantic Web or Web 3.0 [8].

Special Issue Theme

This special issue covers both perspectives of – Web 2.0 and the Semantic Web. In addition to the focus on either of these two technologies, the special issue also covers the "third" approach as well – what *other* technologies contribute to both the Semantic Web and Web 2.0? We are witnessing flourishing of service-oriented architectures, model-driven engineering, and Web-mining technologies, to name but a few, that might have a considerable impact on both Semantic Web and Web 2.0. The special issue tries to answer the following questions. Can these other technologies bridge the controversies between the Semantic Web and Web 2.0, or do they only widen the gap and drive the two approaches further away from each other? Alternatively, can *other* technologies take on the role of matching up with the semantic demands of Web 2.0 applications? Can other technologies help users effectively create, maintain, map between, and use RDF/OWL content, in order to further support Web 2.0 participatory ecosystems of content that is supplied and maintained by their users?

Selected Papers

This special issue brings together eight peer-reviewed papers that represent the current state of the research in the areas of Web 2.0 and the Semantic Web. We grouped the papers into four general sections. The first section covers the topics of collaborative tagging, integration of folksonomies, ontology-based disambiguation of collaborative tags, and novel interaction interfaces for semantically enabled knowledge sharing and grouping. The second section investigates the use of adaptivity and personalization of user interfaces in Web 2.0 and Semantic Web applications. The third section is also related to user interfaces, but from the perspective of traceability and synchronization of two aspects of knowledge representation, one is suitable for machine reasoning and another one is suitable for human use. The final section looks at possible benefits of the combined use of Semantic Web technologies with the techniques of the data mining and model-driven software engineering disciplines in the domains of e-learning and digital libraries.

Tagging and Semantics

The first section of the special issue is dedicated to the topics of collaborative tagging, integration of collaborative tags, and semantic enrichment of collaborative tags. The paper "A system for integration and leveraging of collaborative tags" by Milan Stanković and Jelena Jovanović looks into the problems of integration of collaborative tags created at different locations. Due to the collaborative nature of tagging systems such as del.icio.us, Flickr, and CiteULike, users can easily share content and knowledge. However, once the users move from one collaborative tagging system to another one, the tags are typically encapsulated inside their original systems, while some of those (CiteULike) even do not provide any APIs to access them. To address this problem, Stanković and Jovanović developed the TagFusion system. TagFusion implements different strategies for integration of collaborative tagging systems, such as harvesting tags from all systems a user is subscribed to, and their integration into the tag cloud of the application at hand. TagFusion also supports more advanced usage scenarios where it is possible to automatically tag some content by using the collaborative tags created elsewhere. An important feature of TagFusion is that it distinguishes between human- and machine-created annotations. This can be leveraged in ranking of the discovered resources by giving the higher priority to those resources whose annotator was human.

As the authors of TagFusion state, TagFusion makes one step further toward the idea of the Semantic Web. However, given that different tagging systems are produced by different communities and that they are specific to different contexts, there is a need to consider the integration of collaborative tags by investigating their semantics. This is the problem that Fabian Abel, Nicola Henze, Daniel Krause, and Matthias Kriesell address in their paper entitled "Semantic enhancement of social tagging systems." This paper proposes the GroupMe! system, which combines Web 2.0 and Semantic Web technologies. From the Web 2.0 side, it leverages intuitive user interfaces that allow users to create groups of resources (Web pages, videos, images). Creation of groups, addition of resources to the groups, and any other operation related to the groups are all saved as RDF triples compliant to a set of ontologies that GroupMe! uses. Such an RDF approach to capture group annotations leverages semantic technologies for integration and sharing of groups among the users through the use of Semantic Web benefits. In particular, this eliminates the problems of ambiguity and improves the ranking of the discovered resources.

Adaptability and User Interfaces

Collaborative tagging leverages the idea of collaboration of a number of users on the Web in order to produce shared knowledge (e.g., folksonomies). The key aspect for the success of collaborative technologies, in particular, and Web 2.0 in general, is in the advanced user interfaces that allow users to easily interact with each other

and with the content. While collaboration is widely supported, the main challenge is how to develop techniques for personalization of both Web 2.0 and Semantic Web systems. In her paper "Adaptation and recommendation techniques to improve the quality of annotations and the relevance of resources in Web 2.0 and Semantic Web-based applications," Ilaria Torre recognizes that (semantic) annotation is the major factor for the success of both Web 2.0 and Semantic Web. Therefore, she investigates how different adaptation and recommendation techniques can improve the quality of semantic annotation. Starting from an analysis of weaknesses of the Web 2.0 and Semantic Web approaches, Torre comes up with a set of criteria for improvement of the quality of semantic annotation.

As already mentioned, the success of Web 2.0 is often attributed to the use of advanced user interfaces that provide rich user experience. However, the major challenge is to provide rich-user experience on the Semantic Web. Kay-Uwe Schmidt, Roland Stühmer, Jörg Dörflinger, Tirdad Rahmani, Susan Thomas, and Ljiljana Stojanovic in their paper "Adaptive reactive rich Internet applications" analyze the problem of adaptivity of applications that provide rich user experience. The key challenge is to recognize the current context in which the user is working. To address this challenge, Schmidt and his colleagues propose the concept of Adaptive Reactive Rich Internet Applications. The key idea of the concept is its distinction between offline/design-time and online/run-time levels. At design-time, ontologies are used both to annotate Web applications and conceptually mine Web usage. To enable adaptation, Schmidt et al. propose a lightweight rule language based on the paradigm of reaction rules (event–condition–action). These rules are used on the client-side of Web applications and are triggered as a result of semantically enabled data mining. At run-time, the proposed architecture creates user models on the client-side of Web applications and leverages the created user models as the input of the event processing and rule engine, which is also placed on the client-side of the applications.

Knowledge Representation and User Interfaces

Adaptivity is certainly important for personalization of user interfaces, but an equally important challenge is that of traceability between the machine-processable and human-readable representation of knowledge. Danica Damljanović and Kalina Bontcheva in their paper "Towards enhanced usability of natural language interfaces to knowledge bases" investigate the problem of using natural language as an interface to knowledge bases. Considering this in terms of the Semantic Web, natural language is used as the input representation of user queries. Such queries are automatically translated into formal queries and executed against an ontology and ontology-based repository. Damljanović and Bontcheva survey a number of different systems as per a set of usability criteria, which they also identify in the paper. Based on the conclusions of the survey, they propose a set of recommendations for improving usability of natural language interfaces to ontologies from

the perspective of end users. In this analysis, they included the following aspects: vocabulary restriction, feedback, guided interfaces, personalized vocabulary, and disambiguation strategies.

Usability is a key aspect for the successful document authoring and management. Many different domains have various standards for document and content management (e.g., IEEE Learning Object Metadata for e-learning) along with the accompanying content management tools. However, current practice indicates that very few content authors use these tools in spite of their very advanced features and compliance to standards. The problem is in the usability and habituality of the tools. Namely, content authors stick to the content authoring tools they are familiar with (e.g., Office tools). Similarly, semantic technologies offer many advanced services for document management, but they are typically not well connected with the user-readable representations of documents. Saša Nešić addresses this problem in his paper entitled "Semantic document model to enhance data and knowledge interoperability." This paper presents the Semantic Document Model (formalized in the OWL language), which allows for transforming current documents into so-called semantic documents. Semantic documents are uniquely identified and semantically annotated composite resources, which can be instantiated into human-readable and machine-processable forms. On top of this model, Nešić developed the Semantic Document Management System for managing semantic documents. This system is integrated into Microsoft Office in order for users to be able to make use of semantically enabled services and benefit from the enhancements of the well-known and proven user interfaces for document authoring and management.

Data Mining, Software Engineering, and Semantic Web

Web 2.0 and Semantic Web are not isolated technologies, but they very much make use of the other complementary technologies. In this special issue, we selected two such papers. The first paper authored by Ana Kovačević is entitled "Ontology-based data mining in digital libraries." Data mining is a well-established data management discipline whose major goal is to discover relevant knowledge from (semi-)structured sources of data. As such, it has a very complementary objective to the one of the Semantic Web. In her paper, Kovačević demonstrates how ontologies and data mining techiques complement each other in the domain of digital libaries. Kovačević investigates the problem of the diversity of journal abbreviated names listed in the Journal Citation Reports. The paper illustrates the use of data mining to generate light-weight ontologies of the journal names. The automatically generated ontologies are used in the clustering task of data mining, and the obtained results outperform the results of the clustering task without the use of ontologies.

Current research on the relations between software engineering and the Semantic Web technologies has demonstrated many beneficial synergies [9]. The work of Sonja Radenković, Nenad Krdžavac, and Vladan Devedžić presented in the paper "An assessment system on the Semantic Web" builds on the successful results

in integration between model-driven software engineering and ontology languages [10]. This paper illustrates the use of description logics, underlying formalism of ontology languages, to assess automatically students' assignments where assignments include open-ended questions. The authors make use of the description logic reasoner LoRD, which is fully implemented by using model-driven engineering principles and which is built on top of the recently adopted standard Ontology Definition Metamodel at the Object Management Group. Likewise, the use of model transformations allows the authors to transform the questions and answers represented in the IMS Question and Test Interoperability specification into OWL-based ontology assertions. Once the questions and answers are translated to OWL, ontology reasoning services are used to analyze students' answers.

Summary

As with virtually everything else, one can always find evangelists, devotees, and fans of specific Web technologies. Web 2.0/Social Web and the Semantic Web are no exception to this rule. Still, as Tom Gruber stresses, it is a "popular misconception that the two worlds are alternative, opposing ideologies about how the Web ought to be. Folksonomy vs. ontology. Practical vs. formalistic. Humans vs. machines. This is nonsense, and it is time to embrace a unified view" [2]. Since both Web 2.0 and the Semantic Web have advantages and deficiencies, why not take the best of both worlds and make a synergy of both technologies for the benefit of all users?

In addition, why not identify and tackle problems that neither of the two technologies addresses properly, and make the synergy open for "third-party add-ons"? Note that both Social and Semantic Web lack a more sound software engineering foundation, and both would benefit from deploying advanced, personalized, and multimodal user interfaces for knowledge and data acquisition and sharing. More automation is certainly welcome in the area of semantic annotation, where social tagging and folksonomies represent at best the first step on the ladder. After all, dialog-based human–computer interaction and natural language interfaces are both very social and very semantic-rich, so they can be investigated as a natural extension to the synergy of Semantic and Social Web. Last but not the least, there are still very few applications that really reason over the Web of data. This creates a great challenge for future exploration and integration of the Social Semantic Web with more advanced technologies.

Acknowledgments

This special issue would not have been possible without the significant contributions of many individuals and organizations. Prof. Stefan Voß, the editor of Annals of Information Systems, provided us with invaluable assistance and guidance.

We are also grateful to the reviewers for their dedication in reviewing the papers and providing the authors with substantial feedback. Here we provide the full list of the reviewers:

- Giuseppe Carenini, University of British Columbia, Canada
- Alexandra Cristea, University of Warwick, UK
- Darina Dicheva, Winston-Salem State University, USA
- Juan-Manuel Dodero, University of Cádiz, Spain
- Jon Dron, Athabasca University, Canada
- Adrian Giurca, Brandenburg University of Technology at Cottbus, Germany
- Jelena Jovanović, University of Belgrade, Serbia
- Miloš Kravčík, Open University of The Netherlands, The Netherlands
- Oscar Lin, Athabasca University, Canada
- Sergey Lukichev, Brandenburg University of Technology at Cottbus, Germany
- Alexander Mikroyannidis, University of Leeds, UK
- Michael Minock, University of Umeå, Sweden
- Sasa Nešić, University of Lugano, Switzerland
- Adrian Paschke, Free University Berlin, Germany
- Fernando Sanchez-Figueroa, University of Extremadura, Spain
- Richard Schwerdtfeger, IBM, USA
- Sofia Stamou, University of Patras, Greece
- Heiner Stuckenschmidt, University of Mannheim, Germany
- Carlo Torniai, University of Southern California, USA
- Dunwei Wen, Athabasca University, Canada

We also thank the authors for their efforts in writing and then revising their papers, and we thank Springer for publishing the papers and for a great collaboration throughout the all stages of the work on this special issue.

Belgrade, Serbia — *Vladan Devedžić*
Athabasca, AB, Canada — *Dragan Gašević*

References

1. SW Activity. Semantic Web Activity Statement, W3C [Online], 2008. Available: http://www.w3.org/2001/sw/Activity [Accessed: 2008, December 23].
2. Gruber, T. Collective Knowledge Systems: Where the Social Web Meets the Semantic Web, *Journal of Web Semantics*, 2008, 6(1), 4–13.
3. Mika, P. Ontologies Are Us: A Unified Model of Social Networks and Semantics, *Journal of Web Semantics*, 2007, 5(1), 5–15.
4. Alani, H., Kalfoglou, Y., O'Hara, K., Shadbolt, N. Towards a Killer App for the Semantic Web. In *Proceedings of the 4th International Semantic Web Conference*, 2005, pp. 829–843.
5. Greenberg, J., Robertson, W.D. Semantic Web Construction: An Inquiry of Authors' Views on Collaborative Metadata Generation. In *Proceedings of International Conference on DC Metadata for e-Communities*, 2002, pp. 45–52.

6. Maedche, A., Staab, S. Ontology Learning. In Staab, S., Studer, R. (Eds.) *Handbook on Ontologies, International Handbooks on Information Systems*. Springer, Berlin, 2004, pp. 173–190.
7. Mikroyannidis, A. Toward a Social Semantic Web, *Computer*, 2007, 40(11), 113–115.
8. Hendler, J. Dark Side of the Semantic Web, *IEEE Intelligent Systems*, 2007, 2(1), 2–4.
9. Gašević, D., Kaviani, N., Milanović, M. Ontologies and Software Engineering. In Staab, S., Studer, R. (Eds.) *Handbook on Ontologies*. Springer, Berlin, 2009, 811 p.
10. Gašević, D., Djurić, D, Devedžić, V. *Model Driven Engineering and Ontology Development, 2nd edition*. Springer, Berlin, 2009, 378 p.

Contents

Contributors

Fabian Abel IVS, Semantic Web Group, Leibniz University Hannover, Appelstr. 4, Hannover, Germany, abel@kbs.uni-hannover.de

Kalina Bontcheva University of Sheffield, Department of Computer Science, Regent Court, 211 Portobello Street, Sheffield, UK, K.Bontcheva@dcs.shef.ac.uk

Danica Damljanovic Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield, UK, D.Damljanovic@dcs.shef.ac.uk

Vladan Devedžic FON, School of Business Administration, University of Belgrade, Serbia, devedzic@etf.rs

Jörg Dörflinger SAP Research CEC Karlsruhe, Vincenz-Prießnitz-Straße 1, Karlsruhe, joerg.doerflinger@sap.com

Dragan Gašević School of Computing & Information Systems, Athabasca University, Athabasca AB T9S 3A3, Canada, dragang@athabascau

Nicola Henze IVS, Semantic Web Group, Leibniz University Hannover, Appelstr. 4, Hannover, Germany, henze@kbs.uni-hannover.de

Jelena Jovanović FON, School of Business Administration, University of Belgrade, Jove Ilića 154, Belgrade, Serbia, jeljov@gmail.com

Ana Kovacevic Faculty of Security Studies, University of Belgrade, Gospodara Vučića 50, Beograd, Srbija, fikana1@gmail.com

Daniel Krause IVS, Semantic Web Group, Leibniz University Hannover, Appelstr. 4, Hannover, Germany, krause@kbs.uni-hannover.de

Nenad Krdžavac FON, School of Business Administration, University of Belgrade, Serbia, nenadkr@tesla.rcub.bg.ac.yu

Matthias Kriesell Department of Mathematics, University of Hamburg, Bundesstraße 55, Hamburg, Germany, kriesell@math.uni-hamburg.de

Saša Nešić Faculty of Informatics, University of Lugano, Via G. Buffi 13, Lugano, Switzerland, sasa.nesic@lu.unisi.ch

Sonja Radenkovic FON, School of Business Administration, University of Belgrade, Serbia, sonjam@fon.rs

Tirdad Rahmani SAP Research CEC Karlsruhe, Vincenz-Prießnitz-Straße 1, Karlsruhe, tirdad.rahmani@sap.com

Kay-Uwe Schmidt SAP Research CEC Karlsruhe, Vincenz-Prießnitz-Straße 1, Karlsruhe, kay-uwe.schmidt@sap.com

Milan Stanković FON, School of Business Administration, University of Belgrade, Jove Ilića 154, Belgrade, Serbia, milstan@gmail.com

Ljiljana Stojanovic FZI Forschungszentrum Informatik, Haid-und-Neu-Straße 10-14, Karlsruhe, ljiljana.stojanovic@fzi.de

Roland Stühmer SAP Research CEC Karlsruhe, Vincenz-Prießnitz-Straße 1, Karlsruhe, roland.stuehmer@sap.com

Susan Marie Thomas SAP Research CEC Karlsruhe, Vincenz-Prießnitz-Straße 1, Karlsruhe, susan.marie.thomas@sap.com

Ilaria Torre Department of Computer Sciences, University of Torino, Corso Svizzera 185, Torino, Italy, ilaria.torre@di.unito.it

Section 1: Tagging and Semantics

Chapter 1
TagFusion: A System for Integration and Leveraging of Collaborative Tags

Milan Stanković and Jelena Jovanović

1.1 Introduction

In recent years the way users perceive the Web has changed significantly. From the passive source of information and services, the Web has become, in the eyes of users, a platform for collaboration, a place where their contribution counts. This shift is primarily influenced by the appearance of new kinds of Web sites focusing on socialization, openness, and collaboration that stimulate each individual to participate in enriching the Web content and in the growth of the Web itself. Web sites such as del.icio.us, Flickr, and Facebook belong to this new trend commonly called Social Web (or Web 2.0 in some sources).

The Social Web has offered many new opportunities to Web users, like the ability to easily publish content (using blogs and wikis), share photos and comment on other peoples' photos (using Flickr), or share bookmarks (using del.icio.us). These changes and improvements that the Social Web has brought to Web users should be considered in the context of the original purpose of the Web. In his book *Weaving the Web* [1], Tim Berners-Lee has written: "The web is more a social creation than a technical one. I designed it for a social effect — to help people work together — and not as a technical toy. The ultimate goal of the Web is to support and improve our web-like existence in the world." In this context it is easy to perceive Social Web as a technologically advanced approach to the basic goals of the original Web.

Social Web applications allow users to contribute content to the Web more easily, to publish content even without the knowledge of the underlying technologies (e.g., HTML), to publish opinions without much effort (just by clicking on thumbs up or thumbs down buttons), etc. However, the easiness of contribution and collaboration

Milan Stanković and Jelena Jovanović
FON, School of Business Administration, University of Belgrade, P.O. Box 52, Jove Ilića 154, 11000 Belgrade, Serbia, e-mail: {milstan,jelena}@gmail.com

V. Devedžić and D. Gašević (eds.), *Web 2.0 & Semantic Web*, Annals of Information Systems 6, DOI 10.1007/978-1-4419-1219-0_1,

is not the only significant change that Social Web has brought. The model of use of the early Web where users browsed the content in order to find what they were looking for is getting compromised by the ever larger and increasing quantity of available content. The awareness of the necessity to easily access the right content without browsing is increasing, and it is there where content metadata begins to matter. The ability of information to be easily found is gaining more and more importance on the Web where content is being more and more rapidly added, and that ability is closely related to formally expressed semantics.

To provide advanced services, such as personalization and better use of the available content, machines need formal semantics. To make the Web a place of semantically enriched content relying on ontologies became the mission of a promised Semantic Web. The Semantic Web vision inspired the efforts of research circles to develop and standardize formats for representing semantically rich metadata. Such richly represented metadata would give machines knowledge about content, which would raise their possibilities beyond simple manipulation of data and bring them closer to the possibility to process the content in a more human-like manner. In addition, the idea of semantically rich metadata gave wings to the dream about intelligent software agents capable of performing many actions that are done today by humans [2].

Meanwhile many Social Web sites turned to less precise, but easier to collect form of metadata – tags. Without taking rich knowledge representations much into account, those Web sites have in a short period of time collected huge repositories of tags describing Web resources.

Despite the obvious differences between the formal approach of the Semantic Web and a more pragmatic approach of Social Web, which may lead us to think of them as of opposed directions of tracing the advancement of the current Web, the full potential of the Web may actually arise from merging these two approaches.

To bring tags closer to the needed level of interoperability outlined by the Semantic Web technologies, many approaches have been suggested to enhance their semantic richness [3, 4]. Apart from that, for bringing tagging metadata to its full potential, it is necessary to overcome the current lack of cooperation between tag repositories and find a way to make machines partners in annotation process.

In this paper we address the latter issue, namely the problem of cooperation between repositories of tags as well as the problem of aggregating human annotations with those provided by machines. In Sect. 1.2 we explain basic facts regarding tags and folksonomies. In Sect. 1.3 we go further into analysis of the problem of cooperation between tagging systems and propose a way how this problem could be addressed. We also present a concrete system called TagFusion that implements the proposed approach. Special attention has been given to the possible ways to attract users to use the system, thus making it a valuable metadata repository. In Sect. 1.4 we give an overview of related work and finally give conclusions in Sect. 1.5.

1.2 Background

In this section we explain the concepts of tags and folksonomies, focusing on their advantages and disadvantages, in order to give a solid ground for understanding currently existing problems in the field that we address in this article.

1.2.1 Tagging

Tagging gained significant popularity in Social Web sites, as a way to describe resources (e.g., Web pages, photos, blog articles, etc.) for classification and search. In the tagging process users assign freely chosen keywords to Web resources they wish to describe. This activity is usually incorporated as an optional part of the service offered by a Web site (e.g., on Flickr where adding tags to uploaded photos is optional), but there are also Web sites that relate tagging more closely to their primary service (e.g., del.icio.us where tagging is used to classify bookmarks).

Considerable differences between collaborative tags and taxonomies or folder structures (sometimes used for organizing bookmarks) represent the source of both the popularity of tagging and the drawbacks it brings. Taxonomies are hierarchical and exclusive, thus an object can belong to only one unambiguous category which is in turn within a more general one. Folder trees function similarly. For example, consider a hypothetical user who wants to save an audio file in his music collection. Let it be the recording of Montserrat Caballe performing Puccini's Madam Butterfly. Our user could save this file in the folder c:/music/classical/Puccini or c:/music/classical/Caballe. The reason for this dilemma is obvious: it is often hard for users to assign an asset to one single category. The existence of these two folders would in fact make further searches for this file more difficult by forcing the user to search both locations. On the other hand, tags are neither exclusive nor hierarchical, so if our user could classify the file just by assigning tags he could easily choose tags like classical, Puccini, Caballe, thus making it easy to later search for this file. Apart from the obvious advantage, the chosen set of tags does not give any information to specify that music composed by Puccini and performed by Caballe belongs to classical music genre. When tagging is used, all keywords are considered as equal, regardless of their possible hierarchical or any other semantic relations.

For better understanding of the nature of tags and possibilities for their application, we will look at their main advantages and disadvantages. On the advantages side, we can point out the following:

- Simplicity and low entry barriers: Tagging requires no special skill, and a good will can make any user a successful tagger. No knowledge of predefined vocabularies is required, which makes tagging a very easy and straightforward process for which anyone is eligible.
- Quick adaptation to new terms: Since no predefined vocabularies are used, new terms that appear more and more frequently in IT and other fast developing

domains are easy to emerge and become popular for annotating related resources. Some authors [5] make the analogy with "desire lines," foot-worn paths that sometimes appear in a landscape over time.
- A means for organization of resources: As described in the example with Montserrat Caballe recording, tags provide a way to organize and classify resources, which can be considered superior to traditional classification systems in some aspects.

Apart from the promising list of advantages, there are many limitations and drawbacks that prevent tags from being the right form of metadata for some applications.

- Ambiguity: In tagging systems vocabulary is uncontrolled and there is no way to make sure which tag corresponds to which concept. The word "apple" is the most used and now famous example, since it relates to a fruit, computer hardware manufacturer, and the daughter of Gwyneth Paltrow (who was born at the same time when tagging hype began and – thanks to this coincidence – gained her popularity in research circles).
- Synonyms: Different words or word forms (e.g., plurals) can be used to describe the same concept, and tagging systems provide no means to store the information that two tags relate to the same concept.
- Multiple words and spaces: In some systems (e.g., del.icio.us) users provide tags separated by spaces, and the only way to represent a concept usually described by two or more words is to concatenate them in some way. Different users use different strategies for concatenation (e.g., likeThis, likethis, like-this), and the system ends up with different representations used as a reference to the same concept.
- Basic level problem [6]: When users classify a resource, related terms used for describing it vary from very general to specific. For example, while some passionate drivers could tag photography with the word Audi, for others it would be just a car.
- Lack of semantics: Tags provide limited information about the context in which they are related to the resource being tagged. For example, if a Web page is tagged with the tags "music" and "Madonna" we could not know whether it relates to a page containing some reviews of Madonna's music or it actually contains audio or video samples.

These drawbacks inspire the efforts of scientific community to find solutions that would bridge the gap between the needed level of semantic richness and the level offered by tags [7, 8].

Apart from these inherent disadvantages of tags, there are also considerable problems regarding popular systems for collaborative tagging that should be mentioned.

Tagging systems do not cooperate. For the last couple of years, since tagging systems came in use, many of them (most notably del.icio.us, Flickr, and Technorati) have collected a significant base of annotations, but there is very little effort made to integrate those annotation bases and benefit from quantity of metadata, which hopefully can be used to generate more accurate annotations [9]. Some of the systems

do not provide any possibility to retrieve metadata they have gathered, while others provide some application interfaces (APIs), but neither makes effort to collaborate, nor to facilitate interoperability of the collected tagging data. Some important interoperability issues are discussed in [10].

The other significant problem is that no system can work with both human annotations and automatic annotations provided by, for example, keyword extraction services or autonomous agents. Even systems that do provide interfaces for possible involvement of such artificial users (with del.icio.us API this is theoretically possible) are not able to distinguish automatic annotations from those collected by humans; instead they mix these two types of annotations, thus making them less usable.

1.2.2 Folksonomies

In collaborative tagging systems, benefits of tags as a means of classification are combined with social effects. When multiple users tag a resource with the same tag, we could say that it gains more relevance in the eyes of others. Apart from the quantity of annotations, the confidence in the user who has annotated a resource with a certain tag also plays a significant role in evaluating tag relevance. Consequently, collaborative systems are bringing tagging to the level of an advanced classification scheme. This result of collaborative tagging is generally referred to as folksonomy.

Even though a folksonomy is often (mistakenly) equated with a set of tags created by a certain user community, it actually consists of three sets of entities [11]:

- Users – actors who assign tags to Web resources in collaborative tagging systems
- Tags – keywords assigned by users to resources in order to describe them
- Resources – digital objects being tagged by users (e.g., Web pages, photos, blog articles)

Since in this manuscript we consider folksonomies in a broader context than the one implied by isolated tagging systems, it is necessary also to take into account the origin of tags (i.e., the source system they originate from) as a component that determines the tagging context. Perceived in this way, namely as a combination of the aforementioned components and the source system, tagging context plays the essential role in creating the possibility for improving the semantic richness of tags.

1.2.3 Problem Description

We have already indicated the lack of cooperation between existing collaborative tagging systems as one of the major obstacles for making better usage of available tagging metadata. Many systems like del.icio.us, Flickr, CiteULike,[1] and Technorati

[1] http://www.citeulike.org/