

UNIVERSITÀ DEGLI STUDI DI PISA

**Facoltà di Ingegneria**

**Corso di Laurea Specialistica in Ingegneria Informatica**

**Tesi di Laurea Specialistica**

**Semantic Web gets into social tagging**

Francesco Ronzano

**Relatori**

Prof. Marco Avvenuti

Prof. Alessio Vecchio

Dott. Andrea Marchetti

Ing. Maurizio Tesconi

Anno Accademico 2006/07



# Overview

During the last few years the World Wide Web has experienced continuous and substantial changes in the ways its growing amount of information is produced, organized, accessed and distributed. New paradigms of content creation and sharing and new communication technologies have been introduced in order to fully exploit the potential of this global net of data.

The community of users has been more and more directly involved in these transformations; new possibilities of interaction are continuously enriching the effectiveness of users experience providing new ways to access, organize and interconnect Web information. Moreover the users involvement in the process of Web data access and production has changed. Indeed, the typical user has moved from a passive view of the Web as a great collection of published accessible contents characterized by limited possibility of interaction, to an active one: every user can create and share information communicating with other users and contributing, in different ways, to enrich the great amount of accessible data over the Web. Some important interrelated factors that have greatly contributed to these substantial changes are listed in what follows:

- The growing and capillary diffusion of Internet connections, especially of broadband accesses;
- The progressive shift of a lot of services and activities over the Web, thanks to the advantages obtained in terms of global visibility and availability;
- The increasing number of users acquiring expertise and involvement in the production of Web contents exploiting the most widespread Web technologies, thanks also to the great number of available visual tools and guided procedures.

As a consequence the centrality of Web role in society is more and more increasing and the Web has become a continuously growing pervasive information container characterized by many evolving way to access and exchange data. For instance, we can just consider that the amount of Web information tripled between 2000 and 2003 [HowMuch2003].

In this complex scenario, the need to effectively and efficiently manage this enormous interconnected quantity of data has lead to the introduction of new technological solutions which has mainly affected three distinct but tightly-coupled areas:

1. The integration and interoperability of the globally available informative content;
2. The introduction of new powerful methods to interact with and compose information;
3. The users involvement in the process of content production and their gathering around communities of interest.

These fields of research have respectively given rise to three new specific trends in Web information management:

1. **The Semantic Web;**
2. **The Web 2.0;**
3. **The Social Web.**

In the first part of this document, we analyze each one of these Web approaches describing the supporting motivations of their introduction and their fundamental features. In particular, we explore in more details the Semantic Web because of the completely new point of view that it introduces in the representation of Web information, thus affecting all other Web aspects. We point out its main application fields, giving also some practical example and discussing its future perspectives. Considering the Semantic Web, the Web 2.0 and the Social Web we try to explore their common or complementary aspects attempting to stress their possible synergies. Then we consider also another relevant related question: the diffusion of lexical resources over the Web, focusing our attention on WordNet; lexical resources are increasingly representing an important reference mean to support and better structure the organization of Web contents.

In the second part of this document, we introduce a new possibility to manage and organize information about Web resources: the semantic tagging. Starting from the recent growing diffusion of the tagging activity as a powerful way to socially collect different kinds of descriptive information about Web resources through the association of one or more words called tags, we analyze the possibility to provide semantic support to this task; in this way we foster convergence between the social tagging and the Semantic Web vision. We identify and examine all the fundamental weakness points of actual tagging systems, proving that the introduction of semantics in collaborative tagging can solve or at least reduce a considerable part of them. Indeed semantically tagged Web resources provide a new stronger organizational structure of the information collected through the metadata thus produced. We consider the current obstacles to the availability of an extensible and coherent semantic support to the semantic tagging activity, identifying some way of exploitation of existing Web resources in order to perform this task. We have developed SemKey: it is a semantic tagging system which practically gives the opportunity to exploit the

main advantages due to the introduction of semantics in the tagging process. After providing the description of the architecture and the implementation of our collaborative semantic tagging system, we evaluate its advantages over actual tagging systems in terms of information structuring and retrieval and we point out some relevant suggestions for future works.

