

Leveraging social semantic components in executable environments for learning

Rossana Damiano

Dipartimento di Informatica and CIRMA
Università di Torino - Italy
rossana@di.unito.it

Cristina Gena

Dipartimento di Informatica and CIRMA
Università di Torino - Italy
vincenzo@di.unito.it

Vincenzo Lombardo

Dipartimento di Informatica and CIRMA
Università di Torino - Italy
VRMMP - Torino - Italy
cgena@di.unito.it

Abstract

Learning can benefit from the modern web structure through the convergence of top-down encyclopedic institutional knowledge and bottom-up user-generated annotations. A promising approach to such convergence consists in leveraging the social functionalities in 3.0 executable environments through the recommendation of tags with the mediation of lexical and semantic resources.

This paper addresses such issues through the design and evaluation of a tag recommendation system in a Web 3.0 web portal, “150 Digit”. Designed for schools, “150 Digit” encourages students and teachers to interact with a set of four exhibitions on the historical and social aspects of the Italian unification process in a virtual environment. The web site displays the exhibits and their related documents promoting the users’ active participation through tagging, voting, and commenting the exhibits. Tags become a way for students to create and explore new relations among the site contents, orthogonal to the institutional viewpoint. In this paper, we illustrate the recommendation strategy incorporated in “150 Digit”, which relies on a semantic middleware to mediate between the input expressed by the users through tags and the top-down institutional classification provided by the curators of the exhibitions. Following on, we describe the evaluation process conducted in a real experimental setting, and discuss the evaluation results and their implications for learning environments.

Keywords: tag recommendation, Web 3.0, evaluation, learning environments

1 Introduction

According to W.L. Hosch’s definition, Web 1.0 can be described – using an analogy to file system permissions – as “read-only”, Web 2.0 as “read-write” and Web 3.0 as “read-write-execute”.¹ Following the Web 3.0 principle of executability, the “150 Digit” web portal (<http://www.150digit.it>) has been designed with the goal of creating a virtual environment where schools interact with the exhibitions that celebrate the 150th anniversary of the Unification of Italy. In 150 Digit, a 3D reconstruction allows users to visit the exhibitions; and encourages them to be an active part of the site community by tagging, voting, and commenting the exhibits, and by uploading new contents. The site contents include both the exhibits, with their related documents (such as the curators notes), and user-generated contents, such as multimedia presentations created by teachers and students on the Unification of Italy.

As a result of the user centered methodology on which the site was designed, tagging emerged as a primary issue starting from the de-

sign phase, in the focus groups organized with the teachers involved in the project. Since the pioneering work by [Bateman et al. 2007] tags have been pointed out as an important resource in learning: “The information provided by tags provides insight on learner’s comprehension and activity” (p. 1). Being primarily targeted at educational users, tags play a two-fold role in 150 Digit: on one side the tagging activity, as stressed by the focus groups participants, is part of the educational processes and promotes the linguistic reflection of the students over the site contents; on the other, tags complement the institutional stance on the themes covered by the exhibitions, mirroring the users own understanding and letting new opinions emerge. Also, tagging fosters new correlations of the site contents and can be exploited by the users to navigate in an alternative way to following the paths offered by the site’s information architecture. Finally, user preferences and tags are used to generate recommendations of contents and promote the exploration of the site in a “bottom-up” perspective.

In 150 Digit, the support provided to the users (and educational users in particular) to improve tagging is a recommender module, which was designed and developed in order to meet the project’s specific needs. Suggesting tags to users aims at overcoming the well known trend according to which a site folksonomy stops or slows its growth after some time because the users start to use the same tags, and do not introduce new tags anymore [Trant 2009]. The novelty of the approach developed for 150 Digit is that the tag recommender relies on the semantic description of the exhibitions provided by the curators. Basically, the generation of new tags is obtained through lexical resources, which allow the recommender to expand the meaning of the existing tags, while non relevant tags are filtered out by consulting the semantic description given by the curators. The rationale for this approach is to leverage the exhibitions’ institutional perspective to focalize the generation of new tags and support the teachers work more effectively. At the same time, this strategy aims to avoid the generation of a semantic gap between the institutional categorization of the contents and the folksonomy, keeping them aligned as the latter is expanded. Notice that this can be seen as a variation of the well known “vocabulary problem” (first identified by [Furnas et al. 1987]), i.e. the lack of convergence of terms in user-generated vocabularies.

In this paper we describe and evaluate the tag recommendation system of 150 Digit as a method to improve the contribution of the social semantic components in executable environments designed for learning. The semantic layer of the site relies on a light ontology, WordNet Domains [Bentivogli et al. 2004], a taxonomy of domains originally developed to add semantic information to the meanings of the terms in WordNet [Miller 1995]. In 150 Digit domains are used to categorize the exhibits and the user-contributed contents, providing a background description against which new tags are sought by the recommender. The recommendation of tags exploits the meaning relations encoded in the Italian version of WordNet,

¹<http://www.britannica.com/blogs/2007/07/web-30-the-dreamer-of-the-vine/>

MultiWordNet [Pianta et al. 2002], to expand on the existing tags and propose new ones.

From March to June 2012, students and teachers visited the large exhibition “Fare gli italiani” (“Making Italians”), where they attended a post hoc laboratory in which they were asked to interact with the “150 Digit” portal. These laboratories were the basis for a thorough evaluation of the system. In this paper, we analyze the results of the evaluation, assessing the impact of our approach on the accuracy of the inserted tags, their quantity and typology. Given the data recorded in log files (a sort of indirect observation), we also analyze the users’ behavior in general and compare the folksonomy generated through the system use with a baseline. The paper is structured as follows: after surveying the related work (Section 2), in Section 3 an overview of the “150 Digit” web portal is given, in terms of its design, goals and functionalities, including a preliminary evaluation conducted on the prototype system. Section 4 gives an evaluation of the portal through an experiment with real users. Discussion and conclusions end the paper.

2 Related work

Since the advent of Web 2.0, tagging has attracted much interest in scholars and has been studied under many perspectives. In particular we acknowledge three main areas in the corpus of tag-related research. Tags have been studied with the goal of understanding the behavior and interests of users, letting different tagging styles emerge; more recently, they have been studied as a resource, in an attempt to extract ontologies from user-generated folksonomies; finally, promising attempts have been made to exploit tags in order to provide personalized recommendations and services to users, ranging from tag recommendation to recommendation. In e-learning, tags can be viewed as a means to gain insight on students’ learning [Bateman et al. 2007]; from them, information can be gathered to build user profiles, aimed at making the learning environment adaptive [Ferreira-Satler et al. 2011].

Several attempts have been made to interpret sets of user tags. Users can tag with different purposes: to categorize or describe a resource for future retrieval, or to give an opinion [O’ Donovan 2009]. Concerning the action of tagging in the artwork domain, the results of the experiments in the Steve.museum project [Trant 2006] show that when users add tags on artworks, professional users and non-expert users insert complementary information: non-expert users insert information on the subject of the artwork (such as, in case of a painting, the people and place depicted, the ideas it suggests, the emotions, etc.), while experts provide only “external” information regarding the authors, the historical period, materials and so on. Moreover it emerged that users are generally keen on leaving a trace of what they think and feel.

Other works envisage the complementarity of top-down classification and user-driven classification. [Szomszor et al. 2007] suggest that the best solution for resources accessibility would be to integrate the users’ subjective perspective with traditional classification systems. This could exploit the benefits of both approaches limiting their respective problems. This idea also inspired the work of “150 Digit”, which integrates the curators’ knowledge encoded in the semantic component and the users’ perspective expressed by tags.

Regarding the user participation to both content creation and tag insertion, Nonnecke et al. [Nonnecke et al. 2004] and Preece et al. [Preece et al. 2004] identify the roles of “lurkers” and “posters”, where lurkers are members of online communities who read, but do not post, and posters are the few members who post content. These results have been recently confirmed by Gena et al. [Gena et al.

Accepted for publication.]: the results show that the most participating users contribute in the form of small contributions (clicking on a tag for insertion, clicking on like/dislike) and just a few of them generate bottom-up contents. Analyzing the users’ tagging activity, they reported that 84% of user tags were the ones proposed by the system and just clicked on by the users, while the remaining 16% were inserted by users as free text. This demonstrates that, when available, users tend to select proposed tags instead of inserting new ones, thus providing support to the use of tag recommenders. These findings are also stated in [Ames and Naaman 2007]. The authors reported that in the same domain (photos), users tag more in systems that recommends tags (ZoneTag) than in a system that does not offer tag recommendations (Flickr).

Concerning the recommendation of tags, most approaches rely on statistical techniques (PageRank, evolved into FolkRank [Hotho et al. 2006]) to learn correlations among tags from their co-occurrence in a folksonomy, and use this information to suggest suitable tags for a resource. In our approach, the tag recommendation mechanism relies on the meaning relations encoded in an external resource, i.e., WordNet [Miller 1995] and WordNet Domains [Bentivogli et al. 2004], following the approach proposed by [Xu et al. 2006]. Similarly, [Cantador et al. 2011] proposed a method that uses the YAGO ontology (containing information from Wordnet and Wikipedia) for filtering and classifying tags into a set of purpose-oriented categories (content-based, context-based, subjective, and organizational). The results show that content- and context- based tags are considered superior to subjective and organizational tags in helping a tag recommender component. They found that the transformation of tags into ontology concepts consents inferring semantic relations among concepts for recommendation purposes.

In content-based recommenders, the use of WordNet to improve recommendations is not new. [Degenmis et al. 2007] transform the classic keyword based profiles into semantic user profiles utilizing Wordnet and experienced that semantic user profiles produce more accurate recommendations. [Laniado et al. 2007] propose integrating WordNet in the navigation interface of a folksonomy. In particular using WordNet to build a hierarchy (top-down classification) of related tags (the relatedness is calculated according to well known similarity metrics in Wordnet) can help users navigate and find related resources in del.icio.us. This approach is quite similar to our tag recommendation strategy, in which related tags are suggested on the basis of the hierarchical relation encoded in Wordnet. Finally [Djuana et al. 2011] have found that a backbone ontology, such as the 43 categories incorporated in WordNet, may improve tag recommendation. They automatically learn the ontology from user tags, and use this ontology to improve recommendations by re-ranking the proposed tags on the basis of a collaborative filtering algorithm. They have found that the re-ranking procedure improves precision and recall. We currently do not consider WordNet categories in our recommendation process, though an automatic learning component could be added in order to perform this task.

A strong focus on semantics characterizes the content-based recommendation systems [Pazzani and Billsus 2007], as is the case of 150 Digit. An essential component for content recommenders is a system to describe the items that may be recommended, and this description very often relies on an ontology. For instance, in e-learning, Protus 2.0 tutoring system [Vesin et al. 2012] is a content-based recommender that uses an ontology for knowledge representation and inference engines for reasoning. [Tang et al. 2012] starting from a mining approach combined with fuzzy logic techniques generate a Personal Web Usage Ontology (written in OWL), which enables personalized web resources recommendation. On the side of content-based recommender in the artwork domain we mention

the CHIP artwork recommender². Similarly to 150 Digit, where most content items are constituted by artworks, one of the main goals of CHIP is to demonstrate how Semantic Web and recommendation technologies can be deployed together to improve the access to digital museum collections [Wang et al. 2009]. Results from a user test demonstrate that users prefer content-based recommendations that leverage artwork features, and conclude that domain-specific terms are generally more useful for content-based techniques than generic ones. This finding is in line with our decision to use domain knowledge, encoded in the semantic categorization of the contents provided by the curators, to improve the recommendation of tags.

3 System Overview

The goal of 150 Digit is to provide an open environment where students can visit the exhibitions online and access a wide repository of multimedia items related with the subject of the Unification of Italy. The site contains both institutional contents, taken from the exhibitions, and user-generated contents. Contents can be commented on and tagged by users, thus generating new connections over them. Tags are exploited to group contents on the fly, through a dedicated tag-based search tool; tags and preferences are exploited to recommend contents to the users. By doing so, the site integrates the top-down perspective reflected in the institutional categorisation of content with the bottom-up perspective induced by the users' activity.

3.1 Functionalities and Design

The project encompasses three user profiles: the *editor*, who is in charge of editing and publishing the institutional contents provided by the exhibition curators, and validating the contents uploaded by the students; the profile of the classes, *student* and *teacher*, who can visit the exhibitions, add tags and comments to the exhibits, vote them, and upload new items; and the *registered user*, who does not belong to a class but can visit the exhibitions, vote and tag the exhibits, and create her/his own playlist in a private area. Dedicated tools like the “virtual classroom” (a separate space to comment site contents shared by a group of students under the guidance of one or more teachers) are aimed at improving the quality of the interaction with the site for educational users (for full description of the system, see [Damiano et al. 2011]). Given these profiles, the portal has three main functions: *content management*, *content editing* and *navigation*.

- A content management system allows the site editors to create the site main sections (in 150 Digit, they consist of exhibitions) and categories within these sections, to add contents to the categories and describe them through tags and semantic labels. These labels constitute the semantic layer of the site (as described in Section 3.2).
- A simple content editor lets the educational users edit and publish contents in the existing exhibitions and categories, describing them through tags. During the tagging process, users can ask the system to recommend tags. Suggesting tags to users is a way to contrast the trend according to which a site folksonomy slows its growth because the users stop introducing new tags [Trant 2009]. In 150 Digit, given the educational goals of the site, tag recommendation also serves the purpose of supporting the teachers' work on the linguistic description of the exhibits and documents contained in the site. Web 2.0 functionalities, such as tagging and commenting, are also available as part of the site navigation.

- Site navigation, open to non-registered users, is the same for the three profiles. In addition to the standard navigation enforced by the site's information architecture, users can navigate the contents by following content recommendations or by using the tag-based search tool. In a didactic perspective, exploring the site through the recommendations provided by the site or through the tag-based search can be seen as a way to support the teachers' work in relating the items of the exhibitions into alternative, coherent narratives.

The interaction design of 150 Digit relies on the ‘visit’ metaphor to structure the information. The user can visit the four exhibitions with a standard hypertext-based format, by following the connections over the items induced by tags through the tag-based search, or in a 3D modality (see Figure 1). The portal features a plugin, tested on major browsers, to navigate the exhibitions in a 3D environment, with the aim of making the access to the exhibits more compelling. This approach is borrowed from entertainment (videogames in particular) in order to offer students with an immersive, non textual access modality they are familiar with.

- The standard, hypertext-based navigation follows a classical top-down approach from general categories to detailed information. The information architecture encompasses three layers, namely exhibitions, categories and items; at item level, the user can move across items by following recommendations.
- The 3D navigation contains a set of navigation paths that mirror those experienced by the visitors in the real exhibitions. The use of the same structure in both the standard and the 3D visit is aimed at providing guidance to users in the 3D space. The 3D visit relies on the paradigm of constrained spatial navigation [Burigat and Chittaro 2007], i.e., it is constrained to some fixed positions, in sequential order, where the visitor is “transported” through a stepwise flight simulation (briefly described in [Damiano et al. 2012]).
- The tag-based navigation provides a bottom-up approach to site contents. In this modality, users can take advantage of the search functionalities (by keyword, artwork's title, author, tag, etc.) and sort the items by number of views, users' preferences, and so on.

Users can switch from one modality to another (for instance from 3D to hypertext, from hypertext to tag-based navigation) anytime during navigation, and remain in the same (virtual) location (e.g. the same category or item) after the switch. This approach is intended to stress the parallelism between the various navigation modalities, taking user's need of orientation into account, and giving them the possibility to easily switch among multi-modal information and different viewpoints of navigation.

150 Digit was developed by a multi-disciplinary team, involving AI, computer graphics, interaction design and media experts, and with the participation of the target users in all the phases of the project, from design to prototyping, according to a user-centered, iterative design methodology. The resulting portal integrates different components (social, didactic, informative) in a seamless interface that overcomes the challenges posed by the software integration issues and the content production process. The web 3.0 portal interface design was inspired by usability heuristics and guidelines, as well as by information architecture principles. Moreover the web pages were created in respect of the Italian accessibility law (Stanca Act). A usability expert supervised the interface design together with the web designers, and reported heuristics and guidelines that guided the design decision process.

²<http://www.chip-project.org/>



Figure 1: The visit modalities in 150 Digit. Left, standard hypertext; center, 3D visit; right, tag-based visit.

Different types evaluations were carried out by the project team at different stages of development. In the system design stage, and in particular during the requirement elicitation, a focus group of 5 users, 4 males and 1 females, aged 40-62, was selected. The participants were shown to a set of 15 scenario based static interfaces and the main systems functionalities, labeling and layout with the designers, for 3 hours. In general this group of teachers highlighted the need for textual content to be associated to the exhibits, and for dedicated tools for content creation. They appreciated the proposed interfaces/functionalities, and considered them as valid tools for classroom work and students' involvement. The main findings emerged from the focus group affected the project with changes in both labeling (e.g., "favourites" instead of "playlist") and functionalities. In particular, some of the existing functionalities were modified (for example, teachers suggested to show tag recommendations only on request), and new ones were added: mainly, the possibility of creating a virtual class where students can discuss the exhibits and insert comments that are visible only within this class.

A preliminary evaluation was conducted on a static prototype, which consisted of interface screenshots. This evaluation aimed at verifying the navigation issues (such as breadcrumbs, home button, etc.) in the graphical interface and the users' reception of the social (tagging) and semantic (tag recommendations) functionalities. 5 users were tested. These were teachers, 3 males and 2 females, aged 25-55. The test on the static interface consisted of showing the users screenshots to and discussing the solutions with respect to the aforementioned functionalities, while tag recommendation module test consisted of the accomplishment of a set of tasks, such as tagging or voting an item. The issues that emerged from this evaluation concerned the understanding of the social aspects of the site, such as the role of tags in the fruition of the contents. So, in the re-design, tooltips to explain the meaning of social functions and the possibility of increasing the size of the pictures which illustrate the contents were added.

3.2 Semantic Framework and Tag Recommendation

The need to support prototyping, development and production within a tight time schedule has determined the choice to rely on 'light' semantic tools to leverage the portal recommendation functions. The system semantics rely on WordNet Domains [Bentivogli et al. 2004], a hierarchy of domain labels (169 labels) integrated in MultiWordNet. While most ontologies require expert knowledge to understand their structure (consider for example, foundational ontologies like SUMO [Niles and Pease 2003] or DOLCE [Gangemi et al. 2002]), WordNet Domains lends itself to the use by non expert users, providing an off-the-shelf, portable middleware on the top of which semantic tools can be built.

Semantic Categorization of Contents. In 150 Digit, the recommendations provided by the system rely on a semantic categorization of the contents, with the aim of integrating the social component with the institutional perspective conveyed by the curators in the conceptual organization of the exhibitions. For each exhibition in 150 Digit, each category was associated by the curators to the domains, which, according to them, better describe the category coverage in semantic terms. For example, the "Timeline" category was associated with the "Time_Period" domain, the "Mass media" with multiple domains, "Linguistics", "Photography", "Telecommunication", "Cinema", "Radio", "Telephony" and "Tv". The underlying assumption is that semantic tools (and taxonomies in particular) can provide an effective "external grounding" to the relations over tags, as exemplified by the work of [Markines et al. 2009], that employs taxonomies (such as WordNet) to measure the reliability of the emergent semantic relations among tags in folksonomies, thus providing a sound foundation to the Social Semantic approach.

Tag Expansion and Disambiguation. Differently from standard approaches, which exploit statistical techniques to recommend tags (as in the case of PageRank, re-cast into FolkRank [Hotho et al. 2006]), the tag recommendation mechanism in 150 Digit consists of a constrained expansion of the meaning of existing tags, based on the semantic relations over the lexical items incorporated in WordNet [Miller 1995].

In WordNet, words are gathered into sets of synonyms (i.e. words with same meaning), called *synsets*; synsets are linked according to meaning relations, such as hyperonymy (more general meaning) and hyponymy (more specific meaning). MultiWordNet includes the Italian language and is aligned to WordNet 1.6. The basic *expansion* relies on the synonymy relations among lexical items encoded in synsets:

1. For each user tag, get the corresponding lexical entry from the lemmatizer;
2. Given the lemma, get the synsets from MultiWordNet in which it appears;
3. For each synset found, get all the lemmas contained;
4. Merge the obtained synsets by deleting the repeated entries.

Further expansion relies on querying MultiWordNet for related synsets based on hyperonymy and hyponymy relations at step 3.

The simple expansion mechanism described above however does not guarantee that the recommended tags are actually related to the user tags, due to the polysemy of natural language. In other words, a tag may correspond to more than one lexical entry. To

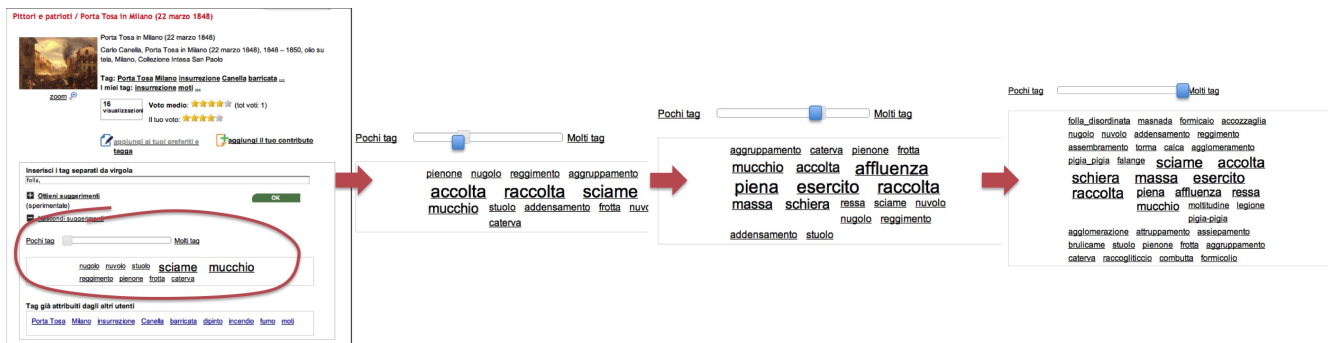


Figure 2: Screenshots of the tag recommendation interface. The slider allows the user to regulate the quantity of recommended tags (from left, “Pochi–Few” tags, to right, “Molti–Many” tags). Recommended tags (here, given the input tag “folla”, i.e., “crowd”) are arranged in a tag cloud. The terms referring to “crowd”, include “mass”, “army”, “bunch”, “swarm”, etc. .

overcome this difficulty, two disambiguation strategies are incorporated in the tag recommender. The disambiguation relies both on ‘syntactic’ knowledge provided by the context of other tags and on the ‘semantic’ knowledge contained in the semantic layer. The ‘syntactic’ disambiguation relies on the context of the other tags associated with the item: for each proposed tag, if it co-occurs in the same synset with one of the context tags, it is included in the recommended tags; otherwise it is discarded. For example, consider the situation in which the tags associated with an exhibit are “emigranti” (emigrant), “pescatore” (fisherman) and “giovane” (youth). Following this strategy, a tag which is a synonym of one of the three tags (for example, “ragazzo”, i.e., young man) will be recommended, while a tag which is not a synonym of any of the three tags (such as “garzone”, i.e. shop boy) will be not recommended. The ‘semantic’ disambiguation relies on the domains attached to the categories, inspired by [Magnini et al. 2002]. Each exhibit inherits the domain labels associated with the category it belongs to (each exhibit belongs to only one category, parallel with the actual arrangement of the exhibition) and with the exhibition itself. These domains provide the semantic context against which the proposed tags are filtered to eliminate the non relevant ones. For example, consider the Italian word “quadro”. This word has two different meaning, “painting” and “control panel”, the first one associated with the “Art” domain in MultiWordNet, the second one associated with the “Electronics” domain. If the disambiguation occurs in the category “Painters and patriots” (associated, among others, with the “Art” domain), only the first meaning of the word “quadro” is considered, while the second one (with its synonyms and other related terms) is discarded because its domains don’t match the category domains.

Interactive Tag Recommendation. In order to let the user control the combination of the expansion and disambiguation techniques described above, the recommendation of tags is accomplished in an interactive fashion (see Fig. 2). If the user enters one or more tags in the system, an auto-completion function shows the possible words given the letters inserted so far; then, the user can then ask the system to propose new, related tags. If no tags have been inserted by the user, the recommendation takes the tags that are already associated with the current item as input (if any, otherwise, the recommendation cannot be made). The amount of recommended tags is regulated by a slider: the user can move the slider from the “Few tags” position (the starting position) to the “Many tags” position, through intermediate positions. Each position corresponds to a different combination of tag expansion and filtering. Figure 2 shows how the cloud of recommend tags grows as the user moves the slider from left (“less tags”) to right (“more tags”), with two intermediate positions between the initial recommendation to the highest expansion

sion of the user inserted tag.

Although the interface allows the user to control the tag expansion mechanism, the presence of hyponyms and hyperonyms may still disorientate them, since their introduction in the set of recommended tags may not be obvious, especially the first time the system is used. In order to overcome this problem, a tag cloud presents the recommended tags, so as to alert the user of the possible presence of unexpected tags. As the user moves the slider, the tag cloud grows or shrinks, and the user can accept one or more of the recommended tags by clicking on them. In the suggested tags cloud, the font size of each tag is given by a combination of two factors: tag frequency in the folksonomy and in the language use. The use of word frequency in language use (taken from a frequency lexicon, “Corpus e Lessico di Frequenza dell’Italiano Scritto”, CoLFIS [Laudanna et al. 1995]) has the function of making more unusual terms less visible in the cloud.

Recommender Architecture. The architecture of the tag recommendation system includes the following components:

- **Lemmaizer:** performs the morphological analysis of the user tag, returning its non flexed form, needed to access the lexical knowledge. For example: “persone” (people), the plural form of “persona” (person) is converted into the singular form. Since most tags are nouns, we chose to consider only the plural to singular conversion. The latter is achieved by using a data base of forms, implemented in MySQL.
- **Expansion Module:** written in PHP, implements the expansion of the user tags along the semantic relations incorporated in MultiWordNet, as described above. Again, MultiWordNet is stored in a MySQL data base and is accessed by a set of PHP APIs.
- **Disambiguation Module:** implements the context–base and the semantic–based disambiguation strategies. This module interacts with the site CMS to get the set of tags that have already been added to the item and the domain labels that are associated with it.
- **Tag Cloud Generator:** this module determines the size of the tags in the generated cloud based on the frequency of tags in the folksonomy and in the lexicon.

Item recommendation.

The content recommendation function relies on two complementary approaches: a collaborative filtering approach [Schafer et al. 2007] and a semantic approach. So, the user is presented with two sets of

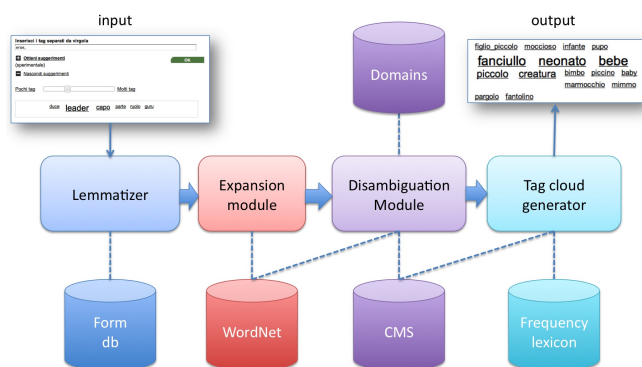


Figure 3: *The semantic architecture of the web 3.0 portal.*

recommendations, one of which is based on the preferences given by the other users, and the other is based on the tags added to the items.

The semantic-based recommendation selects the items to recommend based on the shared tags with the current item. Items are ranked according to the number of tags they have in common with the given item. Items with the same ranking are re-ranked according to the category to which they belong: items from the same category (and the same exhibitions) as the current item are preferred.

The social recommendation is based on the preferences expressed by the community of the users, and is inspired by the technique of collaborative filtering [Xu et al. 2006; Sarwar et al. 2001]:

1. Given the current content, select its highest vote;
2. Select all the users who have given the same vote to that item;
3. For each of these users, select the items to which the user has given the same (or higher) vote ; If the set is empty, set the vote to vote - 1;
4. Rank the selected items by their highest votes;
5. Select the first n contents;

The user is presented with the two sets of recommendations (tag-based and preference-based); the difference between the two is communicated by different labels, “150 Digit recommends” and “Other schools recommend ” respectively. In case the same item appears in the two sets, the duplication is eliminated.

3.3 Preliminary Evaluation

Given the logs of the first six months of publication of the web site, a preliminary evaluation of the users’ acceptance of the site functionalities was conducted, and of the recommender system in particular. Only front-end users were considered for social functionalities (content generation and tagging), while for the semantic analysis of tags back-office users were also taken into consideration, as they benefit from this kind of recommendation because they are requested to tag exhibits as part of the publication process

During the first six months, 347 users logged onto the site, 149 (42.93%) active teachers, 199 (57.35%) regular visitors. Of the regular visitors, 61 users (31%) were associated to classes. It is important to note that teachers and classes were explicitly contacted by the committee in order to promote the portal. These teachers/classes were randomly selected over a set of teachers/classes

who regularly participate in trials organized by the Ministry of Education. This evaluation, conducted for prototype refinement and experiment design, was split into two parts: the first concerning the generic Web 2.0, i.e. social, functionalities and the second relating to the 3.0, i.e. executable, functionalities.

Web 2.0 functionalities.

The users’ behavior in relation to social functionalities can be analyzed in terms of:

- their participation in content creation,
- their tagging activity,
- the quantity and the typology of inserted tags.

With regards to the uploading of user-generated content, 11 virtual classes of the 51 registered classes (21.57%) inserted new contents (a total of 29 new contents, while the institutional contents are 271). In detail, 3 classes with same teacher inserted more than half of contents (51.72%), one class inserted 13.79% of contents and another one inserted 10.34% of contents. Thus 5 classes out of 11 (45.45%) inserted 76% of contents.

In total 404 tags (duplicates included), either freely inserted by users or selected among the tags suggested by the system, have been collected since the beginning of the experimentation. More specifically 297 tags (73.51%) were proposed by the system and just clicked on, while the remaining 107 (26.41%) were inserted by users in free text. 28 users out of 347 (8.07%) inserted tags. Of these 17 were teachers working with their virtual classes (61%) and 11 were regular visitors (39%). In particular a teacher using the site both as a regular visitor and with her 5 classes inserted almost half the tags (201 tags, 49.75%). Note that this teacher, and her classes, were the same that inserted more than half the contents. Another teacher both as visitor and with her class inserted 16.83% of tags (namely 68 tags), while another class created an above average number of tags (31 tags, 7.67%). The remaining classes (31.71%) inserted an average of 5.9 tags per class, while regular visitors (32.14%) inserted an average of 5 tags per user. The low user participation in both content creation and tag insertion confirms the results of [Nonnecke et al. 2004] and [Preece et al. 2004] and replicates the dichotomy between “lurkers” and “posters” mentioned above.

Not all the tagged contents in exhibition received the same number of tags, see Table 1. The frequency of tags with respect to exhibitions/sections needs to be balanced with the number of contents present in each exhibition/section. In general, the number of tags is proportional to the number of content items in the exhibitions, with two exceptions: “La bella Italia” and the “Extra contents” sections. “La bella Italia” received the least user attention in term of tags, despite its relevant number of contents; the “Extra contents” section, i.e. the section containing schoolgenerated contents that are not strictly related to the main exhibitions, was particularly successful. This success is not surprising, as the contents generated by classes receive more attention by the classes themselves. Moreover, the insertion of a new content implies the insertion of tags.

It is interesting to compare the number of tags received by each exhibition with the number of visits of the real and virtual exhibitions. In the real world, the exhibition “Fare gli italiani” had the highest number of visitors, followed by “La bella Italia”, “Stazione futuro”, and “Il futuro nelle mani”. The number of visits received by the exhibitions on the web site “Fare gli italiani” is still the most visited exhibition followed by “La bella Italia”, “Il futuro nelle mani”, and “Stazione futuro” (the last two being almost on a par). While the



Figure 4: Content recommendation in 150 Digit. On the left, the tag-based recommendations (“The systems recommends you”); on the right, the preference-based recommendations (“Schools recommend you”).

Exhibition/Section	Number of tags	Number of contents
“Fare gli italiani”	179 tags (44.31%)	135 (45%)
“Extra contents”	70 tags (17.33%)	28 (9.33%)
“Il futuro nelle mani”	37 tags (9.16%)	48 (16%)
“Stazione futuro”	33 tags (7.92%)	30 (10%)
“La bella Italia”	16 tags (3.96%)	38 (12.6%)
“The places (of current exhibitions)”	16 tags (3.96%)	9 (%3)

Table 1: The most tagged exhibitions/sections

trend regarding the former exhibitions has also been confirmed by the 150 Digit taggers activity, the latter ones, in particular “La bella Italia”, reveal a much lower number of tags with respect to their virtual visits. An explanation could be that the sections/exhibitions receiving more tags are those whose contents are more pertinent to the topics covered by the study programme. Regarding the tagged contents, 109 items have been tagged, with an average number of 3.71 tags per item. However the distribution of tag per item is not homogeneous. 10 items (9.17%) received a number of tags more than twice the average, as detailed in Table 2. The other items (90.83%) received a number of tags ranging from 1 to 7. More specifically 4 items (3.67%) received 7 tags, 5 items (4.59%) received 6 tags, 10 items (9.17%) received 5 tags, 7 items (6.42%) received 4 tags, 17 items (15.60%) received 3 tags, 32 items (29.36%) received 2 tags, 24 items (32.02%) received 1 tag. Notice that most of the items received a low number of tags. The most used tags (“history”, “unification”, “Italy”, “risorgimento”, “tradition”, etc) reflect the historical context of the web site contents (the celebration for the 150th anniversary of the unification of Italy), while the others reflect the artwork content (“woman”, “women”, “food”), namely subject related tags. The remaining 254 tags have been used with these frequency values: 198 tags (49%) have been used once, 47 tags (11.63%) have been used twice, 9 tags have been used 3 times. To sum up we can conclude that a few tags are used more than once, while most of the tags are used once, or twice at most. However these considerations must also take into account the limited sample of users involved in the trial.

From these data, we concluded that the users’ behavior with the 150 Digit system is coherent with the social functionalities largely reported in the literature. Different groups (less active and more active users) emerge for the quantity of uploaded contents and added tags, with the distribution of tags featuring the most common tags, in line with the themes of the exhibitions.

Web 3.0 Functionalities

The data set collected in the 6-month testing of the prototype system was employed to conduct a preliminary evaluation of the adequacy of our approach to the recommendation of tags. Our working hypothesis is that, if the approach is correct, the semantics of the folksonomy should, to some degree, match the institutional categorization of contents, thanks to the use of the semantic layer in

the recommendation of tags. In order to obtain a semantic description of the folksonomy, comparable with the description of the categories in the semantic layer, we adopted the very same resource employed to encode the semantic layer itself, i.e., WordNet Domains (see Section 3.2). To do so, tags were associated with the domains of the corresponding terms in MultiWordNet, thus obtaining a picture (though a coarse one) of the semantics of the folksonomy. The obtained representation can be straightforwardly compared to the description of the categories encoded in the semantic layer, which were entered with the relevant domains by the curators.

As a preliminary step of the analysis, duplicated tags were eliminated from the folksonomy; then each tag was associated to the corresponding lemma in MultiWordNet and all the synsets in which the lemma appears were collected. Following this, the relevant domains for each tag were retrieved through the synset ids (in MultiWordNet, domains are associated with synsets).

With this mapping, the distribution of domain labels in the folksonomy were investigated and compared with the institutional labels, both at site and category level. In order to compare the domains associated with the folksonomy with the institutional domains associated with the website categories the folksonomy domains were ranked according to their frequency. Table 3.3 gives the ranking of domains according to their frequency of association to the user tags at site level. As expected, the top domain is Factotum (44.97%), the domain assigned to lemmas having a generic meaning. Beside other generic domains, such as “Quality” (5.35%) or “Person” (5.27%), most of the domains in this rank, such as “Geography” (5.46%), “Military” (4.24%), “Politics” (4.04%) and Art (3.66%) are highly relevant to the themes dealt with by the exhibitions, which narrate Italy’s complex evolution after its unification in terms of political, social and cultural aspects. So, if “Buildings” (5.64%) can be related to the monuments and buildings that appear as pictures or location in many exhibits, “Administration” (4.21%) refers to the institutions and administrative regions mentioned all along the narration of Italy’s historical evolution.

In order to gather more insight, we extended the semantic comparison of tags and institutional domains to the highest level of detail of the categories (each exhibition includes several categories). However, since most data in the tagset concerned the exhibition “Fare gli Italiani”, we limited the category-level comparison to this exhi-

Table 2: Most tagged items

Item	Exhibition/Section	Number of tags
“Adunata”	“Fare gli italiani”	34 tags (8.42%)
“Calendario 2011”	“Fare gli italiani”	16 tags (3.96%)*
“Blog 150 anni insieme”	“Fare gli italiani”	12 tags (2.97%)*
“Cibo per le feste”	“Extra contents”	12 tags (2.97%)*
“L’Italia, Selargius, la Sardegna nei 150 anni”	“extra contents”	12 tags (2.97%)*
“Carabinieri in alta uniforme e Corazziere”	“Il futuro nelle mani”	10 tags (2.48%)
“Il pane, simbolo dell’unita”	“Stazione futuro”	9 tags (2.23%)
“Foggia e l’Unita’ d’Italia”	“Extra contents”	8 tags (1.98%)*
“La classe....di una volta”	“Fare gli italiani”	8 tags (1.98%)*
“Rete-Mondo Moda”	“The places”	8 tags (1.98%)*

Rank	Domain	Hit number
1	Factotum	1092 (44.97%)
2	Buildings	137 (5.64%)
3	Geography	133 (5.46%)
4	Quality	130 (5.35%)
5	Person	128 (5.27%)
6	Military	103 (4.24%)
7	Administration	102 (4.21%)
8	Politics	98 (4.04%)
9	Art	89 (3.66%)
10	Sociology	84 (3.45%)

Table 3: Ranking of domains according to the matching tags.

bition. “Fare gli Italiani” contains 20 categories, with an average of about 3.68 labels for each category. By comparing the domains associated with each category to those associated with tags added to contents of the same category, we found a significant overlap. The comparison showed that the average overlap between folksonomy domains and institutional domains is 61.02%, that is, 61.02% of the domains extracted from the folksonomy of a certain category matched the institutional domains in the corresponding category. More interestingly, for each category the topmost ranked domain in the folksonomy (the domain that was associated with most tags in the category) always matches one of the institutional domains. For example, the topmost ranked domain in the folksonomy for the category “The Migrations” is “Geography”: this domain, together with “Sociology”, was associated by the curator with that category.

The evaluation methodology described has two obvious limitations. Firstly, the limited coverage of the folksonomy by MultiWordNet: of the 1957 tags in the folksonomy, only 865 can be found in MultiWordNet (44.2%). Secondly, we did not perform any kind of disambiguation of tags, so the representation of the folksonomy in terms of domains may reflect the ambiguity of tags. Notwithstanding these limitations, we considered the overlap of the institutional and the folksonomy domains satisfactory, so the tag recommendation strategy was maintained, extending it to work with no text input by the user: in the current version of the website, the input to the tag recommender can be given explicitly, by inserting one or more tags, or implicitly, by requesting the systems to suggest tags based on the tags associated with the current item.

4 Experimental Evaluation

During a few months, students and teachers visited the large and permanent exhibition “Fare gli italiani” (“Making Italians”), and then attended one of the three post hoc laboratories where they were asked to interact with the 150 Digit portal. These laboratories constituted the basis for a thorough evaluation of the system

Laboratory	Control	Experimental
On the trail of migrants	11	10
The suitcase of the historian	2	3
Making Italians E-book:	2	3

Table 4: Laboratory sessions attended, with numbers of the control and experimental groups, respectively.

with the goal of assessing the acceptance of the tag recommender and the effectiveness of the Web 3.0 (a Social Semantic Web approach, borrowing the words of [Markines et al. 2009]) over the Web 2.0 approach in tag recommendation. During the lab sessions, the students were requested to analyze the items related to the exhibits and to create and upload new materials, also adding tags in the meantime. In the experimental group, the students interacted with the system, receiving tag and content recommendations, while for the control group, the recommendation module was disabled in order to gather the control data. Notice that the users in the control group could not use the tag recommender, but were able to see the other users tags (in the style of Web 2.0), when adding new tags to the contents.

In the following, we analyze the results, evaluating the impact of our approach on

- the tagging activity of users, i.e. the quantity and the typology of the inserted tags;
- the semantics they convey.

The latter point implies an analysis of the tag sets generated by the two versions of the system: one is completely user-generated, while the other one mixes user-generated tags and semantic recommendations.

4.1 Tag Analysis: Methodology and Results

Hypothesis. We hypothesized that the tag set generated by the experimental group would be quantitatively and qualitatively different from the tag set generated by the control group. In particular, the folksonomy of the experimental group, having benefited from the tag recommender, should be larger and more heterogeneous.

Design. The first group of students (control group) interacted with a modified version of the system, without the semantic recommender (independent variable). In this phase subjects received only social recommendation, namely the recommended tags that are the most used tags. The second group (experimental group) of students interacted with the regular version of the system, with the semantic recommender enabled.

Participants. 15 classes (9 secondary school classes, and 6 high

school classes) for the control phase with 33 registered users³ on the web site vs. a total of 298 users participating to the laboratories. 16 classes (10 secondary school classes, and 6 high school classes) for the experimental phase with 42 registered users vs. a total of 255 users participating to the laboratories.

Apparatus and Materials. The laboratory rooms were equipped with a set of computers (Windows-based PCs) and one Interactive whiteboard (mainly for the lab conductor). Users browsed the web site using MS Explorer 8. The performances were traced by means of an ad hoc logging system. Users were given written instructions.

Procedure. The classes used the system during one of the following laboratories:

- **On the trail of migrants:** Through multimedia workstations students faced the journey of Italian emigrants of the early twentieth century: the baggage, the passenger list, landing at Ellis Island and the interrogation. Every student group took the role of a (emigrant) character. Then, with a leap through time, they relived the journey of new migrants from Guinea, Afghanistan, Kurdistan and other countries different from Italy. At the end of the laboratory students had to tag the characters on 150 Digit;
- **The historian's suitcase:** Before visiting the exhibition, by means of 150 Digit, students were introduced to the narrative choices made by the historians that curate the exhibition. They then chose an exhibition category and looked for its historical sources. At the end of the lab they uploaded (on 150 Digit) and tagged a document containing their work;
- **Making Italians E-book:** In groups students were assigned specific roles to make observations, collect data and produce images along the way. After the visit, in 150 Digit, each group, created a humorous and original page of text and animated images using Prezi editor, on the themes of the exhibition or on how it was received. The e-book was then uploaded into the system and tagged.

Table 4 summarizes which laboratories were attended during the control and the experimental phase. Each class was free to attend their preferred laboratory. In the first month of experimentation each class interacted with the modified version of the system, without the semantic recommender (control group). During the second month of experimentation each class interacted with the regular version of the system, with the semantic recommender enabled (experimental group). In both conditions, students were required to analyze the items related to the exhibits and to create and upload new materials, adding tags while doing both these operations. Users in the experimental group were asked to insert at least one of the suggested tags.

It is important to note that users interacted with the system in groups. Every class was split into 2-4 groups. So the number of users registered into the system was less than the total number of real users attending the laboratories.

Control group results. The 33 users of the control group inserted a total of 133 tags, with an average of 4.03 tags per registered user, and 8.06 tags per class. 10 users (24.24%) inserted almost half the tags (49.62%), in this way confirming the presence of active users in a community that are more active than others in content contributions [Nonnecke et al. 2004]. The same consideration applies to classes: 4 classes (26.67%) inserted more than half the tags (55.64%). Table 5 shows the most used tags, namely all the tags

Tags	Frequency	Percentage
Cosa Nostra	6	4.51 %
mafia	6	4.51 %
assault	5	3.76 %
migration	5	3.76 %
war	5	3.76 %
emigration	4	3.01 %
Clash	3	2.26 %
collective act	3	2.26 %
education	3	2.26 %
Gold Rush	3	2.26 %
immigration	3	2.26 %
maid	3	2.26 %
civil war	2	1.50 %
conflict	2	1.50 %
fight	2	1.50 %
massacre	2	1.50 %
murder	2	1.50 %
Naples	2	1.50 %
presentation	2	1.50 %
Public Schools	2	1.50 %
school	2	1.50 %
Topic	2	1.50 %
trench	2	1.50 %

Table 5: Most used tags - Control Group

Categories	Frequency	Percentage
Migrations	46	34.59%
Mafias	21	15.79%
The First World War	18	13.53%
The school	15	11.28%
The power of the unity	10	7.52%
The Second World War	9	6.77%
The massmedia	5	3.76%
The Church	3	2.26%
The campaigns	3	2.26%
The consumption	1	0.75%
Italy cities	1	0.75%
Painters and Patriots (2011)	1	0.75%

Table 6: Most tagged categories - Control Group

used more than once. Notice that i) the two most used tags are synonyms (i.e. “Cosa Nostra” and “mafia”); ii) the other top-ranked tags, i.e. migration-emigration-immigration, and assault-war-clash, share the same meaning; iii) these mentioned tags reflect the content of the most tagged categories, namely “Migrations”, “Mafia”, and “The First World War”; iv) finally, a great number of tags (62, which is 46.62% of the total number of tags) were used only once, which shows great sparsity in the use of tags. The control group classes uploaded 20 new contents in total (5 new contents per class). Only 4 classes uploaded new contents since the “On the trail of migrants” laboratory does not require users to upload material.

Table 6 shows the most tagged categories, which clearly reflect the themes of the most attended laboratories.

Experimental group results. The 42 registered users of the experimental group inserted in total 214 tags, with an average of 5.09 tags per user, and 13.37 tags per class. 8 users (19.05%) inserted almost half the tags (49.07%), thus confirming also in this case the presence of active users in a community [Nonnecke et al. 2004]. The same consideration applies to classes: 5 classes (31.25%) inserted more than half the tags (54.87%). Table 7 gives the most used tags, namely all the tags used more than twice.

³Users interact with the system in group. Every class was split into 2-4 groups. Thus the number of registered users on the systems was less than the total number of users attending the laboratories.

Tags	Frequency	Percentage
farmer	8	3.74%
worker	6	2.80%
merchant	5	2.34%
illiterate	4	1.87%
Migrant	4	1.87%
Cameo	3	1.40%
maid	3	1.40%
miner	3	1.40%
mother	3	1.40%
poor	3	1.40%
shopkeeper	3	1.40%
unemployed	3	1.40%
Veteran	3	1.40%
well-off	3	1.40%
woman	3	1.40%
young	3	1.40%

Table 7: Most used tags - Experimental Group

The tags used once are 108 (50.47%), while tags used twice are 10.75%. Thus most tags (more than 60%) are re-used very little or not at all. Concerning the most used tags, i.e. farmer, worker, merchant, illiterate, migrant, etc., we should notice these are mainly the tags reflecting the subjects presented in the “Migration” category.

The experimental group classes uploaded in total 21 new contents (3.5 new contents per class). Notice that just 6 classes uploaded new contents, since the “On the trail of migrants” laboratory does not require to upload material. Table 8 shows the most tagged categories, which reflect the themes of the most frequented laboratories.

For what concerns the slider that regulates the number of recommended tags, its use was analyzed in terms of the positions of the cursor selected by users when they accepted the recommendations. The hypothesis was that an even distribution of the cursor positions would confirm the users’ understanding and acceptance of the slider:

- Tags from position 1 were selected 84 times (39.25%): in this position, the expansion strategy considers the tighter semantic relations encoded in MultiWordNet, i.e., synonymy, and the candidate tags are disambiguated against the domains attached to the current category;
- Tags from position 2 were selected 54 times (25.23%): in this position, the expansion is extended to the hyponymy and hyperonymy relations, but the disambiguation is also extended to take the context of the existing tags into account;
- Tags from position 3 were selected 40 times (18.69%): in this position, the domain-based disambiguation is removed;
- Tags from position 4 were selected 36 times (16.82%): in this position, both disambiguation methods are removed;

As shown by the usage of the slider tool, users seem to accept and understand its usage correctly, since all the cursor’s positions on the slider were employed, with an obvious prevalence of the initial position.

4.2 Tag Recommender Evaluation

Using the same approach described in Section 3.3, the semantic representation of the two tag sets (one generated by the control group and the other by the experimental group), was compared and given

Categories	Frequency	Percentage
Migrations	150	70.09%
The Futurist	17	7.94%
The New Officine	11	5.14%
The mafia	6	2.80%
Campaigns	6	2.80%
Unification of Italy	4	1.87%
The First World War	4	1.87%
It began with their	4	1.87%
Consumption	3	1.40%
Gallery of shops	3	1.40%
Painters and Patriots (2011)	2	0.93%
Officine Grandi Repairs	2	0.93%
The Second World War	2	0.93%

Table 8: Most tagged categories- Experimental Group

in terms of the domains associated with the tags they contain. For each tag set, we computed the overlap of its domains with the institutional domains, category by category (since the association with domains is at category level, i.e., each category was associated by the curators with a set of domains).

For each tag, we collected the domains to which it is associated (using the synset-domain mapping encoded in MultiWordNet). For each category of the exhibition, we obtained a set of domains, (the folksonomy domains) ranked according to the number of tags to which each domain is associated⁴. Each set of ranked domains constitutes a rough semantic representation of the tag set in terms of the taxonomy of domains encoded in MultiWordNet. Tables 10 and 9 report the top ten domain labels for each tag set. The control group tag set refers to 44 different domains; the experimental group tag set refers to 55 different domains. Each domain is accompanied by the number of times it is associated with a tag in the folksonomy (“Hit number” in the tables). For each set, some tags could not be employed for this evaluation because they are not present in MultiWordNet (for example, because they are proper nouns, neologisms, etc.) or because they don’t have a domain label associated in MultiWordNet.

The two sets of ranked domains were compared. First of all, in order to assess if the two sets of domains (obtained in the two experimental sessions) were significantly different from the statistical point of view, the χ^2 test was calculated. The statistic shows that the difference in the two distribution are significant ($\chi^2(67)=108.54$, $p<0.001$).

A more thorough comparison shows that the two sets have 29 domains in common; also, it was observed that only 4 of the 29 shared domains are in the ten topmost ranked domains of both sets (“Sociology”, “Psychological features”, “Military”, “Person”).

By extending the comparison to the categories, the χ^2 test shows that only the different frequencies of distribution of domains belonging to “Migrations” and “The First World War” are significant, obtaining respectively $\chi^2(40)=91.39$, $p<0.001$, and $\chi^2(13)=52.60$, $p<0.001$.

Secondly, the overlap of each set of folksonomy domains with the set of institutional domains was measured. Similarly to what emerged from the dataset collected during the system tuning phase (described in Section 3.3), the top ranked domains in the folksonomy overlap with the institutional domain labels. Moreover, by observing the overlap at category level, we found that, in both tag

⁴Again, notice that the same tag can be associated to different domains due to its intrinsic polysemy (i.e., a tag belonging to more than one synset) or because a synset is associated to more domains.

Domain	Hit number
person	113 (21.86%)
commerce	21 (4.06%)
military	21 (4.06%)
agriculture	17 (3.29%)
sociology	17 (3.29%)
psychological_features	15 (2.90%)
quality	15 (2.90%)
biology	13 (2.51%)
medicine	12 (2.32%)
animals	10 (1.94%)

Table 9: Top ranked domain labels in experimental group.

sets, the topmost ranked domains match at least one of the institutional domains assigned to that category, notwithstanding the low number of domains associated with the categories by the curators (an average 3.68 domains per category). For example, in the “Migrations” category, the domain in common between the folksonomy and the institutional domains is “Sociology” (17 hits in the control group and 10 in the experimental group) where the institutional domains of the category are “Geography” and “Sociology”); for the “The Second World War” category, the most frequent domain (9 hits in the control group and 6 in the control group) is “Military”, which matches the institutional domains associated with that category (“Diplomacy”, “History”, “Military”). The only exceptions are given by some categories for which less than 5 tags were collected. To summarize, although the two distributions of domains are statistically different, they are not significantly different in terms of their overlap with the institutional domain labels. However, it is necessary to specify that a full comparison is impossible. Due to the uneven distribution of tags in the category and to the partial coverage by MultiWordNet, the comparison was possible for only 6 categories out of 35; in all other cases, the tags of the users could either not be found in MultiWord-Net or there were no tags at all in that category (for one of the two tag sets).

The analysis of the semantics of the folksonomy conducted through the domains confirms the prototype analysis findings (Section 3.3), i.e., that the folksonomy domains significantly overlap with the institutional domains. In summary, given the domain based analysis, we can conclude that the assumption on which the recommendation strategy relies, that the institutional domains match the semantics of the site categories as perceived by the users, is confirmed by the data.

As for the effectiveness of the recommendation strategy with respect to the baseline (no recommendation at all), no significant differences in the overlap of folksonomy domains with institutional domains were found between the control and the experimental group; i.e., the overlap does not grow using the recommender. In this way the tag recommendation can seem ineffective. However, the limited size of the tag sets collected during the experimentation makes the comparison difficult, and although a statistical difference does exist, it cannot be traced back to the overlap with the institutional domains. On the other hand, this result can be considered a success in an educational context, because tag insertion in focused trials driven by teachers leads students to select the tags in a very accurate and focused way (with and without the recommenders), and this may have contributed to reducing the difference between the two situations.

5 Discussion

This section summarizes the results of the evaluation and analyzes the tag sets of the experimental and control groups in further de-

Domain	Hit number
sociology	52 (13.07%)
military	37 (9.3%)
person	34 (8.54%)
history	23 (5.78%)
school	17 (4.27%)
law	15 (3.77%)
pedagogy	14 (3.52%)
psychological_features	10 (2.51%)
time_period	8 (2.01%)
linguistics	7 (1.76%)

Table 10: Top ranked domain labels in control group.

tail, with the goal of explaining the difference between the two sets emerged from the domain-based analysis.

As can be expected, users in the experimental group inserted more tags than users in the control group (5.09% tags vs. 4.03% tags per user), showing that the use of the recommender helps users find more tags. Also, users in the experimental group tended to insert more new tags compared to the control group: there are 85 (64%) distinct tags in the control group tag set, and 147 (69%) distinct tags in the experimental group tag set. This finding is positive for the evaluation of the tag recommender, because it shows that its use can increase the variety of tags. However, this is not clear-cut: since users were requested to insert tags at the end of each laboratory session and to reason about which tag to insert (e.g., they were explicitly asked to tag the discussed artworks and the uploaded material), the number and the variety of inserted tags may have been biased by these given instructions, blurring the differences between the two groups. In addition to this, the laboratory context and the presence of the teachers may have led students to be more accurate in tagging.

Concerning the emergent semantics of the generated folksonomy, the number of domains referred to by the tags in the experimental tag set is higher than the number of domains referred to by the tags in the control tag set (55 domains vs. 44 domains). Even if these results do not show a clear benefit in the use of the tag recommender (again, also due to the structure of the experimental context), the number and distribution of tags in the experimental group seem to have been positively affected by the use of the tag recommender.

To better understand the meaning of the tags inserted by the users, further semantics may be required. In order to gain more insight, two more investigations were conducted: first, the tags were analyzed according to their lexical types, trying to register the difference in the two tag sets; second, the distribution of tags in the WordNet Domains in the two tag sets were compared, trying to relate it to the use of the tag recommender.

To study the distribution of tags based on their lexical types, the collected tags were clustered into WordNet categories. A semi-automatic procedure⁵ performed a WordNet dictionary lookup to obtain the top-level categories that could be deduced from the correspondence with the lexicographer file organization⁶. Only tags contained in the WordNet dictionary were mapped to WordNet categories. The obtained classification for control and experimental group is detailed in Table 11.

⁵Disambiguation was performed manually. Each times a tag belongs to more than one category the right meaning was manually checked by confronting the tag and tagged item.

⁶<http://wordnet.princeton.edu/man/lexnames.5WN.html>

Lexnames	Experimental group	Control group
adj.all	31	2
adj.pert	1	0
compound	7	6
invented	12	0
noun.act	19	45
noun.animal	4	0
noun.artifact	14	7
noun.attribute	2	1
noun.cognition	2	10
noun.communication	3	5
noun.event	3	8
noun.feeling	7	0
noun.food	1	1
noun.group	6	18
noun.location	7	10
noun.object	0	1
noun.person	77	15
noun.plant	0	1
noun.relation	1	0
noun.state	12	0
noun.substance	2	1
noun.time	0	1
NULL	1	1
verb.stative	2	0

Table 11: Comparison between Wordnet lexnames

The χ^2 test was calculated in order to assess if the frequency of distribution of tags obtained in the two experimental session were significant. The statistics show that the difference in the two distributions is significant ($\chi^2(23)=125.47$, $p<0.001$). What seems to emerge from this classification and from the comparison of the results of the two experimental conditions is that the presence of the semantic recommender led users to tag adjectives frequently (14.5%) and nouns denoting people (noun.person: 36%), such as farmer, merchant, worker, etc very frequently. The use of these words may be partly explained by the topics covered in “On the trail of migrants” lab, which focuses on people. However, this was the most frequented laboratory, as well as the “Migrations” category, in both groups (not only in the experimental one), and the most used tags in the control group do not reflect this trend. Indeed, students in the control group often used nouns denoting acts or actions (noun.act: 33.8%), and nouns denoting groups of people or objects (noun.groups: 13.51%) more than nouns denoting people (noun.person: 11.28%). From this observation we can conclude that the semantic recommendations of tags promote a more precise description of the subject represented in the artwork. This is demonstrated by the wide use of tags belonging to the Wordnet person category. This finding also confirms the results obtained in the Steve.museum project [Trant 2006], according to which users prefer tags related to the subject of the artwork and to the description of what is represented in it.

As illustrated in the previous section, the comparison of the two tag sets conducted by analyzing their implicit semantics through Multi-WordNet Domains confirmed the statistical difference between the two sets, implying that the use of the tag recommender actually affects the tagging behavior of users, although it does not prove that the overlap of the folksonomy is increased with the use of the recommender. In particular, we tried to relate the distribution of domains in the tag set of the experimental group with the functioning of the tag recommender. Remember that, by using a slider, the users can affect the behavior of the recommender, reducing or increasing the number of recommended tags (see Section 3.2). In the starting position, the recommendation strategy is more restric-

tive, as it applies different disambiguation techniques, which are progressively removed as the slider is moved to the following positions. In parallel, the search for recommended tags is extended from tighter semantic relations (synonymy) to looser ones (hyponymy and hyperonymy). As a result, in the the last slider position, no disambiguation is performed and all available semantic relations are considered.

By analyzing the distributions of domain labels that emerged from the two sets of tags (control and experimental group), we note that the difference resides mainly in two facts:

- In the experimental group, when the recommender was in use, there were 26 more domains (such as “Exchange”, “Administration”, etc.) than in the control group (there are 16 shared domains in the two distributions). These domains tend to be more specific than the shared ones;
- In the experimental group, the most abstract domains in the hierarchy have more tags related to them than the control group. For example, “Person” moves from 34 related tags in the control group tag set to 113 related tags in the tag set of the experimental group.

These two complementary findings show that the growth observed in the experimental tag set is related both with the diversity of domains it contained (with more specific domains added) and with the distribution of the tags in the domains, with more of the tags associated with the generic domains. The users generated more tags with the help of the recommender, and this increase affected both the generic and the specific domains, which increased in number due to the focalization brought by the use of the recommender.

These findings are consistent with the use of the tag recommendation slider. The increased number of specific domains can be attributed to the 39.25% of times the users accepted the recommended tag with the position “A” in the slider: in this position, tags are constrained to the institutional domains, which are generally quite specific, such as “Telecommunications” or “Transport”. The higher number of tags related with generic and abstract domains (such as “Psychological features”, or “Person”) are consistent with the fact that the remaining positions of the slider expand the existing tags along the hyperonymy relations encoded in MultiWordNet, in which 16.82% cases where no disambiguation was conducted (with the slider in D position, where the highest number of recommended tags is generated by the system).

Last, but not least, a difference between the two tag sets lies in the domain/tag relationship. For example, if we consider the most tagged category, “Le migrazioni” (Migrations), we observe that in the tag set generated by the experimental group, tags and categories tend to increase proportionally. Notwithstanding this growth, the set of domains emerging from this tag set remain consistent with actual semantics of the category: for example, consider the domain “metrology”, brought in by the tag related with the distances covered by the migrants, or “religion”, which is related to the description of the ethnic groups involved in the migrations.

Given the analysis conducted, we can conclude that a mixed approach, which relies on leveraging the curatorial knowledge (encoded in semantic format) to support social functions such as tagging, is suitable to the development of active learning environments. The use of the tag recommender contributes to keeping the tagging activity of the educational users aligned with the categorization of contents provided by the curators, at the same time sustaining the growth of the obtained folksonomy and its focalization in terms of variety. By promoting the linguistic reflection on the use of tags, the recommender helps the teacher focalize the students’ work with the beneficial input of the curatorial knowledge.

6 Conclusions

In this paper we described “150 Digit”, a web portal on the 150th anniversary of the Unification of Italy, which has been designed and implemented with a Social Semantic Web approach in mind. The 150 Digit portal displays the contents of the exhibitions organized in 2011 for the anniversary, and encourages the active participation of the educational users through social functionalities such as tagging, voting, and commenting, and through the creation and upload of new contents. A 3D visit contributes to making the environment more compelling for students.

In particular, we described the semantic-based tag recommendations incorporated in “150 Digit”. The users’ tagging activity introduces new connections over the site contents, expanding and complementing the categorization of the exhibitions provided by the curators. The tag recommender exploits the semantic description of the curatorial categorization of contents to sustain the growth of the folksonomy, keeping it aligned with the institutional perspective, with benefits for the focalization of the activities in the schoolwork.

After describing the design of “150 Digit”, mainly based on an iterative, user-centered methodology, we illustrated the evaluation conducted of the tag recommendation function, discussing its results and possible benefits for learning environments.

Given the users’ activity logs collected for 150 Digit, which include tagging, commenting, uploading and saving content, current methodologies (for example, see [Ferreira-Satler et al. 2011]) consent the use of this information to generate more adaptive recommendations. As future work, we envisage the improvement of the tag recommender by extracting user profiles from the users’ behavior.

Acknowledgments

This work was partially supported by the project “150digit. L’Italia delle scuole” funded by Comitato Italia 150. The authors would like to thank: Comitato 150 and Esperienza Italia 150 (especially (Claudia Cugnasco and Marina Bertiglia), CSL – Università di Firenze; Indire – MIUR, Firenze; CIRMA – Università di Torino. This work was carried out with Virtual Reality & Multi Media Park S.p.A. (Fabrizio Nunnari, Marco Squeo, Shanti May, Davide Di Giannantonio). We also thank the users who took part in the focus group and in the tests and the anonymous reviewers for their help and advice.

7 Authors

Rossana Damiano, PhD, is an Assistant Professor at the Department of Computer Science of the University of Torino, Italy. Her interdisciplinary research activity is centred on intelligent applications for cultural heritage, ranging from interactive multimedia systems to applications of virtual agents. In particular, she is interested in the role of semantic models in new media production and applications.

Vincenzo Lombardo, PhD, is an Associate Professor of Informatics at the University of Torino, Italy. He works on models and applications of knowledge-based system to multimedia communication. He carries on a production activity in multimedia production, hosted by events at the international level.

Cristina Gena is an Assistant Professor at the Department of Computer Science of the University of Torino, working in the area of intelligent user interfaces. She completed her Ph.D. in Communication Science (University of Torino) in 2003, with a thesis on the evaluation of user-adaptive systems. Her current research activities address user modeling, adaptive web systems and their evaluation,

context-aware systems, semantic web, web 2.0, usability and interaction design. Her contribution is based on experiences gained within these fields.

References

- AMES, M., AND NAAMAN, M. 2007. Why we tag: motivations for annotation in mobile and online media. In *CHI*, ACM, M. B. Rosson and D. J. Gilmore, Eds., 971–980.
- BATEMAN, S., BROOKS, C., MCCALLA, G., AND BRUSILOVSKY, P. 2007. Applying collaborative tagging to e-learning. In *Proceedings of the Workshop on Tagging and Metadata for Social Information Organization (WWW’07)*.
- BENTIVOGLI, L., FORNER, P., MAGNINI, B., AND Pianta, E. 2004. Revising the wordnet domains hierarchy: semantics, coverage and balancing. In *Proceedings of the Workshop on Multilingual Linguistic Resources*, Association for Computational Linguistics, 101–108.
- BURIGAT, S., AND CHITTARO, L. 2007. Navigation in 3d virtual environments: Effects of user experience and location-pointing navigation aids. *International Journal of Human-Computer Studies* 65, 11, 945–958.
- CANTADOR, I., KONSTAS, I., AND JOSE, J. M. 2011. Categorising social tags to improve folksonomy-based recommendations. *J. Web Sem.* 9, 1, 1–15.
- DAMIANO, R., GENA, C., LOMBARDO, V., NUNNARI, F., SUPPINI, A., AND CREVOLA, A. 2011. 150 digit. integrating 3d visit and social functions into a web 3.0 learning-oriented approach. In *Broadband and Wireless Computing, Communication and Applications (BWCCA), 2011 International Conference on*, IEEE, 136–143.
- DAMIANO, R., LOMBARDO, V., GENA, C., AND NUNNARI, F. 2012. Guidance for web 3d in cultural heritage dissemination. In *Proceedings of the 17th International Conference on 3D Web Technology*, ACM, 186–186.
- DEGEMMIS, M., LOPS, P., AND SEMERARO, G. 2007. A content-collaborative recommender that exploits wordnet-based user profiles for neighborhood formation. *User Model. User Adapt. Interact.* 17, 3, 217–255.
- DJUANA, E., XU, Y., LI, Y., AND JOSANG, A. 2011. Ontology learning from user tagging for tag recommendation making. In *Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Volume 03*, IEEE Computer Society, Washington, DC, USA, WI-IAT ’11, 310–313.
- FERREIRA-SATLER, M., ROMERO, F., MENENDEZ-DOMINGUEZ, V., ZAPATA, A., AND PRIETO, M. 2011. Fuzzy ontologies-based user profiles applied to enhance e-learning activities. *Soft Computing-A Fusion of Foundations, Methodologies and Applications*, 1–13.
- FURNAS, G., LANDAUER, T., GOMEZ, L., AND DUMAIS, S. 1987. The vocabulary problem in human-system communication. *Communications of the ACM* 30, 11, 964–971.
- GANGEMI, A., GUARINO, N., MASOLO, C., OLTRAMARI, A., AND SCHNEIDER, L. 2002. Sweetening ontologies with dolce. *Knowledge engineering and knowledge management: Ontologies and the semantic Web*, 223–233.

- GENA, C., CENA, F., VERNERO, F., AND GRILLO, P. Accepted for publication. The evaluation of a social adaptive web site for cultural events. *User Model. User-Adapt. Interact.*
- HOTH, A., JÄSCHKE, R., SCHMITZ, C., AND STUMME, G. 2006. FolkRank : A ranking algorithm for folksonomies. In *LWA, University of Hildesheim, Institute of Computer Science, K.-D. Althoff and M. Schaaf, Eds., vol. 1/2006 of Hildesheimer Informatik-Berichte*, 111–114.
- LANIADO, D., EYNARD, D., AND COLOMBETTI, M. 2007. A semantic tool to support navigation in a folksonomy. In *Proceedings of the eighteenth conference on Hypertext and hypermedia*, ACM, New York, NY, USA, HT '07, 153–154.
- LAUDANNA, A., THORNTON, A., BROWN, G., BURANI, C., AND MARCONI, L. 1995. Un corpus dell'italiano scritto contemporaneo dalla parte del ricevente. *III giornate internazionali di analisi statistica dei dati testuali 1*, 103–109.
- MAGNINI, B., STRAPPARAVA, C., PEZZULO, G., AND GLIOZZO, A. 2002. The role of domain information in word sense disambiguation. *Natural Language Engineering* 8, 04, 359–373.
- MARKINES, B., CATTUTO, C., MENCZER, F., BENZ, D., HOTH, A., AND STUMME, G. 2009. Evaluating similarity measures for emergent semantics of social tagging. In *Proceedings of the 18th international conference on World wide web, April*, Citeseer, 20–24.
- MILLER, G. 1995. Wordnet: a lexical database for english. *Communications of the ACM* 38, 11, 39–41.
- NILES, I., AND PEASE, A. 2003. Mapping WordNet to the SUMO ontology. In *Proceedings of the IEEE International Knowledge Engineering conference*, 23–26.
- NONNECKE, B., PREECE, J., AND ANDREWS, D. 2004. What lurkers and posters think of each other. *Hawaii International Conference on System Sciences* 7, 70195a.
- O' DONOVAN, J. 2009. Capturing trust in social web applications. In *Computing with Social Trust*, J. Golbeck, Ed., Human-Computer Interaction Series. Springer London, 213–257.
- PAZZANI, M. J., AND BILLSUS, D. 2007. Content-based recommendation systems. In *The Adaptive Web*, 325–341.
- PIANTA, E., BENTIVOGLI, L., AND GIRARDI, C. 2002. Developing an aligned multilingual database. In *Proc. 1st Intl Conference on Global WordNet*.
- PREECE, J., NONNECKE, B., AND ANDREWS, D. 2004. The top five reasons for lurking: improving community experiences for everyone. *Computers in Human Behavior* 20, 2 (March), 201–223.
- SARWAR, B., KARYPIS, G., KONSTAN, J., AND REIDL, J. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, ACM, 285–295.
- SCHAFER, J., FRANKOWSKI, D., HERLOCKER, J., AND SEN, S. 2007. Collaborative filtering recommender systems. *The adaptive web*, 291–324.
- SZOMSZOR, M., CATTUTO, C., ALANI, H., O'HARA, K., BALDASSARRI, A., LORETO, V., AND SERVEDIO, V. D. 2007. Folksonomies, the semantic web, and movie recommendation. In *Proceedings of the Fourth European Semantic Web Conference*, 71–85.
- TANG, J., HUI, S., ZHOU, B., FONG, A. C. M., AND HONG, G. 2012. Generation of personalized ontology based on consumer emotion and behavior analysis. *IEEE Transactions on Affective Computing* 3, 2, 152–164.
- TRANT, J. 2006. Exploring the potential for social tagging and folksonomy in art museums: Proof of concept. *New Review of Hypermedia and Multimedia* 12, 1, 83–105.
- TRANT, J. 2009. Studying social tagging and folksonomy: A review and framework. *Journal of Digital Information* 10, 1.
- VESIN, B., IVANOVIĆ, M., KLAJNA-MILIĆEVIĆ, A., AND BUDIMAC, Z. 2012. Protus 2.0: Ontology-based semantic recommendation in programming tutoring system. *Expert Systems with Applications* 39, 15, 12229 – 12246.
- WANG, Y., STASH, N., AROYO, L., HOLLINK, L., AND SCHREIBER, G. 2009. Semantic relations for content-based recommendations. In *K-CAP*, 209–210.
- XU, Z., FU, Y., MAO, J., AND SU, D. 2006. Towards the semantic web: Collaborative tag suggestions. In *Collaborative web tagging workshop at WWW2006, Edinburgh, Scotland*, Citeseer.