Semantic Annotation and Tagging in the KP-LAB environment: mining the learning and knowledge dynamics.

A. Locoro¹, A.M. Scapolla¹, D. Grignani¹

¹ DIBE, Biophysical and Electronic Engineering Department, University of Genoa, Via All’Opera Pia 11/A, 16145 Genova, Italy, {angela.locoro, scapolla, daniele.grignani}@unige.it

This paper introduces a system, named “Knowledge Practice Environment”, which focuses on innovative practices in knowledge creation and evolution. Design and implementation are based on the “triological” approach, an innovative pedagogical theory which aims at modelling collaborative knowledge-intensive activities around shared artefacts. In particular, this paper describes the design and development of the Semantic Annotation tool and the Tagging tool, which provide functionalities to create named relationships between object of activities, sticking comments to them and annotating them by free text tags, as well as tags from structured vocabularies. The amount of both ontology-driven and non-structured semantic data, collected during the process of explicating knowledge, allows to capture models of social intelligence behaviour and to outline the dual nature of shared categorization and subjectivity in constructing meaningful activities. Suggestions on how to frame and to analyze data with this dual focus are provided. Some applications of such analysis are discussed as a preliminary study both from the system and the user perspectives. Emergent research topics on annotations and tagging activities and knowledge mining are introduced as evidence of the potential and benefits of this approach.

Keywords Knowledge Practices; Semantic Annotation; Collaborative Tagging; Knowledge Mining

1. The KP-LAB project and environment

KP-Lab [1] is an Integrated Project sponsored by the 6th EU Framework Programme in the Information Society Technologies, Technology-Enhanced Learning program, and aims at creating a learning system that facilitates innovative practices of sharing, creating and working with knowledge in education and workplaces. It is coordinated by the University of Helsinki and the project’s partners, 22 organizations, come from research communities that are active both in the pedagogical and technological domains. The project promotes co-evolution of individual and organizational learning with technology through the development of a learning system based on technological, theoretical, pedagogical, and social innovations [2].

The KP-Lab learning system provides the participants tools for reflecting on, making visible, and transforming their knowledge practices. In order to be truly productive, collaborative technologies cannot be fully specified beforehand but need to co-evolve with social practices and be further modified according to the users’ emerging needs and practical innovations. Consequently, the KP-Lab is really a “laboratory”, a test bed for developing transformative knowledge practices for educational institutions and workplaces.

The main characteristics of the learning environment, can be summarized as follows:

- shared working spaces to organize activities around shared objects and to interact at personal and community levels;
- support to organize the community and to structure the learning process;
- support to reflective activities on the shared objects and the learning context, e.g. through commenting, semantically annotating, creating conceptual maps, data mining services...;
- awareness services to trace the knowledge evolution process that is embedded in the practices of the members of a community.

1.1 The Knowledge Practise Environment

The Knowledge Practice Environment (KPE) consists of the set of tools that users can access from a web portal. It is possible to view the existing work environments, named Shared Spaces (SSP), and to create and to configure new ones using the KPE tools. The possibility of customizing the SSPs is larger than it is in traditional learning management systems. Instead of following predefined “course models”, users can choose how to arrange and to structure their working spaces. The teacher maintains the role of organizing and guiding the process but students can play a primary role in the working space. They can add material and define relations between the shared objects, they can initiate discussions on each object and create conceptual maps of the space they are working at. An exhaustive presentation of the KPE is available in [3] and in [4]. The former focuses on technical aspects of the KPE, while the latter presents the environment from a pedagogical point of view and details its functionalities.

In synthesis the main KPE tools concern:
SSP management: to view the existing SSPs as a network or as a list, and the links created between them and to create a new SSP.

Process planning: to organize the work within the SSP by tasks and milestones and to assign resources, users and timelines to each phase of the work.

Resources management within a SSP: to create, organize and visualize various kinds of resources (file uploaded, web links, wiki pages, notes), to share resources with other users, to manage vocabularies, generic like the Project Based Learning (PBL) vocabulary or created specifically in the proper domain of the SSP.

Commenting and linking resources: to create named relationships between all the objects present in the SSP, and to comment each object with the possibility of tracing different threads of comment (Fig. 1, a) - b))

Semantic tagging: to annotate resources with tags/concepts from the vocabularies of the SSP and to provide helpful suggestions based on the analysis of the SSP objects’ content (Fig.1, c) - d)).

Semantic search: to search the SSP objects on the basis of the tagging operations and of the output of similarity search services.

Collaborative semantic modeling: to define visual model languages and visual model of the SSP resources.

Process logging and analysis tool: to trace and provide data about the work ongoing within the SSP.

Meeting support tool: to save conversations as written/graphical resources in the SSP.

Fig. 1 a) A slice of the KPE GUI: arrows with labels represent named relationships between objects while tiny notes with a circled number depicts the number of total comments attached to each knowledge artefact. b) Opening the comments generates the comment threads view where people can read, create a new thread or reply to existing comments. c) A list of tags from an existing structured vocabulary is presented to the user. d) An example of tags assigned to objects.

1.2 Semantic Annotation and Tagging in KPE

The KPE system relies on the ontological representation of the shared space and its knowledge artefacts, as well as the users’ actions, which include commenting and tagging activities. The system model is defined in the TLO
(Trialogical Learning Ontology) as an OWL (Ontology Web Language) Schema [5] and data are stored using the RDF (Resource Description Framework) format [6] as triple statements (subject, predicate, object). Users management is modelled using the FOAF [9] ontology. The semantic engine, called SWKM (Semantic Web Knowledge Middleware) [7], provides all the services and functionalities to store, retrieve and manage all the system data.

The Knowledge Annotator tool is part of the KPE system and provides functionalities to conduct collaborative tagging on the basis of structured SKOS (Simple Knowledge Organization System) [8] vocabularies. A free tags vocabulary is also available in each SSP. It contains free text terms that users can create during the ongoing phase of knowledge elicitation. Semantic annotations activities include also comments of SSP objects structured as argument threads. The creation of named relationships between pairs of artefacts can be conducted by either adding a hierarchical label (“is_a”) or terms from a link vocabulary as well as free text labels.

2. Mining the knowledge process

In the field of organizational knowledge creation one of the main models is that of [10], where a continuous interaction and rich interchange between tacit and explicit knowledge describe the evolution of the community expertise. What are the peculiarities of both tacit and explicit knowledge? How can technology support and to what extent machines can incorporate this knowledge creation and evolution modality?

In educational or professional scenarios users are involved in the organization of the knowledge material according to personal understanding and classification attitude, while sharing ideas and coordinate actions with others. Tacit knowledge has more to do with the individual subjectivity and intuitions layered in the cognition abilities of everyone. Individual knowledge constantly feed collective one, through an iterative elaboration process of ambiguities and redundancies until an explicit common background conceptualization is reached.

Explicit knowledge is the final result of this process and can be formalized in a model of shared environments that represents information across the community, unifying its identity and purposes.

Actions taken on problem solving, learning tasks and knowledge practices are based on this dual nature of communication and semantics.

Studies of the most popular web based systems for tagging contents presented in [11] and [12] claim the potentialities of these practices in capturing social networks dynamics. In such researches the process of adding tags to contents is framed according to its characteristics, potential evaluation and analysis of motivation and re-use.

The following sections are dedicated to the description of how the KPE tools deal with this kind of knowledge creation models. In particular the tagging and commenting activities are analyzed and evaluated as crucial ones to handle both individual and social contributions.

2.1 Making tacit knowledge explicit: learning domains ontologies from tags

In knowledge intensive environments, such as KPE, users concentrate their efforts on the management of shared knowledge objects.

The main goals of associating keywords to artefacts are giving a descriptive label to contents as well as depicting their main features. The process of reasoning on analogies and differences among a group of objects is characterized from the term discovery of synonyms and opposites. Moreover, the owner or responsible of a resource needs to be identified. Evolutions of artefacts according to time spans have to be traced. Selective tasks such as synthesis and discourse structures have to be produced. Categorization activities have also a major role in knowledge sharing processes.

The treatment of the descriptive labels associated to the artefacts is able to provide semantic services, which have to manage the dimensions of ambiguities and redundancies arising during collective knowledge elicitation. Most of the collected information of informal tagging is affected by the so called polysemy, where the same word may have different meanings, and synonymy, where different terms stand for the same referent. This semantic heterogeneity is then reinforced by the disparate levels of abstraction, specification, view points emerging from each individual’s style of classification and perspective. On the other hand collective dynamics, mediated by the same culture, language and interests inside a community, may result to be a strong mechanism of convergence until a shared understanding is reached.

Once the ongoing phase of meaning negotiation arise to a final disambiguation step and an agreement is found, the new piece of knowledge can be used to evolve and refine the system knowledge [13].

The level of emerging semantic resulting from this activity is a straightforward way to obtain concepts able to complement the static part of the well formed system knowledge with a more dynamic one. This can help to maintain domain ontologies, to let them evolve and change over the community supervision.
To this purpose semantic services inside the KPE use to some extent ontology learning techniques [14] from tags in order to provide search services and automatic content classification. Suggestions to the users based on the most popular tags provided by automatic frequency counts are also supported. This is an early step towards the use of knowledge mining services for a more accurate retrieval of information and system service customization through semantic enriching.

2.2 Users, knowledge artefacts, tags and relationships: analysis dimensions.

The KPE system provides different views of the data. This reflects the need to visualize the data according to some main dimensions such as the “Community View”, which focuses on the social network interactions within a SSP, the “Content View”, showing the artefacts connected through named relationships and tagged with vocabulary terms as well as free text comments, the “Process View”, which outlines the temporal dimension of tasks and procedures, and the “Tailored View”, providing a customized selection of activities and objects in the system.

Further to this an export tool has been designed to extract data based on pre-defined analysis scenarios and, at this early stage, is able to aggregate data based on:

- the “User activities”, summing up the number of different types of objects created by each user;
- the “Content items”, providing a schema of qualitative and quantitative metadata and properties on contents;
- the “Tasks”, which has the same purposes of the content items analysis but on a different object type;
- the “Social networks analysis”, depicting tables of associations such as those with users as creators of objects and other users as actors over the same objects;
- the “User actions”, which are extracted from the system logs and are able to capture single and atomic operations that a user does on an object (creation, modification, opening, deletion and so on).

A discussion has started among the project partners on how to evolve such analysis towards innovative fields of pedagogical researches. To this purpose a unified frame of actors, objects and annotations models is a new analysis dimension that is going to be explored. Some directions on how to analyse such models are suggested in [15] and [16].

According to this approach, measures of associations can be mined from users and tags, objects and tags, objects and relationships, tags and relationships. These measures are for example similarities within users based on their vocabulary usage, similarities of tags within the same content, classification schemes based on associations within contents related through named links, metadata discovery of objects and processes based on the combination of both tagging and linking activities.

Knowledge patterns can be extracted and analyzed according to semantic overlapping dynamics across communities of practice working inside the system.

2.3 Comments on knowledge artefacts

Comments are a way of expressing opinions, posing questions, discussing different positions and viewpoints and even influencing the directions of a discourse. Comments are an important type of users-produced contents and represent the most informal and dynamic part of the discourse that users carry out on knowledge artefacts.

Analysis of comments based on classification and text mining techniques can be used in order to extract interactions within subjects and between subjects and objects of the commenting activity. Natural Language Processing approaches are moving towards evidencing attitudes and opinions in text, as such kind of information has become of great usefulness for business, sociology and psycholinguistic studies [17].

A classification of comments in KPE can start, for example, from mining subjectivity as well as polarity computation of negative and positive sentences with the scope of discovering what is the prevailing opinion on a particular topic. Interpersonal distance dynamics can also be extracted in order to cluster data according to sub-communities of aggregation.

This can help teachers as they could receive a feedback of the ongoing activities as well as a feeling of the classroom progressions and dynamics.

On the other hand, extraction of the main topics of the discourse as well as argumentative structures can be used to feed the system semantics.

Mining technologies for detecting critical patterns and concept maps extraction using comments have been already designed and adopted in KPE [18] but viable proposals here investigated are still new and provide an original contribution to the work in progress.

3. Conclusions and Future Work
This paper has presented the KPE knowledge annotation tools as a means to tag, link and comment knowledge artefacts inside the SSP application. Some suggestions on exploiting free text annotations and ontology-driven labels have been given according to state-of-the-art of mining technologies.

Emerging researches on social networks systems are considered and contextualized in the knowledge creation and evaluation perspectives. Suggestions on new mixed analysis models of social networks, objects and annotations are taken into consideration.

Our future work is going in the direction of doing experiments on free text and unstructured data, taking advantage of knowledge mining methods as well as of the combination of structured and unstructured information in order to obtain meaningful models of collaborative knowledge creation dynamics.

We think this can be considered as an original contribution to the discussion on emerging analysis dimensions which aims at a deeper capture of the social intelligence behaviour in technology based learning environments and re-use of such information to further improve systems and tools in charge of assisting knowledge practices.

Acknowledgements The work of the authors is supported by the KP-LAB project funded by the 6th EU Framework Programme in the Information Society Technologies, Technology-Enhanced Learning program. We thank all colleagues and partners involved in the design, implementation and evaluation of the environment.

References

[4] M. Lakkala, S. Paavola, K. Kosonen, H. Muukkonen, M. Bauters, H. Markkanen. Main functionalities of the Knowledge Practices Environment (KPE) affording knowledge creation practices in education, 2008 (Submitted for publication, will provide precise references with the final version of the paper).
[8] http://www.w3.org/2004/02/skos/
[18] D5.6: Specifications for the Knowledge Matchmaker (V.2.0), the Knowledge Synthesizer (V.1.0) and the Analytical and Knowledge Mining Services (V.1.0), Deliverable 5.6 of KP-LAB Project, 2009, to be published.