### Project Deliverable Report

**D6.1 – Social and informal learning support design**

<table>
<thead>
<tr>
<th>Work Package</th>
<th>WP6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task</td>
<td>6.1, 6.2</td>
</tr>
<tr>
<td>Date of delivery</td>
<td><strong>Contractual:</strong> 01-11-2008  <strong>Actual:</strong> 01-12-2008</td>
</tr>
<tr>
<td>Code name</td>
<td>D6.1  <strong>Version:</strong> 1.0</td>
</tr>
<tr>
<td>Type of deliverable</td>
<td>Report</td>
</tr>
<tr>
<td>Security</td>
<td>PU</td>
</tr>
</tbody>
</table>

| Contributors | Paola Monachesi (UU), Eelco Mossel (UU), Stefan Trausan-Matu (UPB-NCIT), Vlad Posea (UPB-NCIT), Kiril Simov (IPP-BAS), Petya Osenova (IPP-BAS), Aneliya Tincheva (IPP-BAS), Laska Laskova (IPP-BAS) |
| Authors (Partner) | Paola Monachesi (UU), Eelco Mossel (UU), Vlad Posea (UPB-NCIT), Kiril Simov (IPP-BAS), Petya Osenova (IPP-BAS), Aneliya Tincheva (IPP-BAS), Laska Laskova (IPP-BAS) |
| Contact Person | Paola Monachesi (UU) |
| WP/Task responsible | Paola Monachesi (UU) |
| EC Project Officer | Mr. M. Májek |

### Abstract (for dissemination)

The goal of this Workpackage is to develop services that facilitate learners and tutors in accessing formal and informal knowledge sources in the context of a learning task. More specifically, we will support the learner through a personalized search. A Common Semantic Framework will be developed which will provide recommendations on the basis of the user profile, his interests, his preferences, his network and obviously the learning task. An advantage of this system with respect to a standard Google search is that the list of search results will be prioritized and categorized according to the conditions specified by the learners. In addition, a trust dimension will be added since the output proposed will take into account the opinions of the user's trusted network of friends.

### Keywords List

- Social media applications
- Ontologies
- Tagging
- Social networks
- Semantic desktop
- Conceptual annotation
# Table of Contents

Executive Summary ......................................................................................................... 3  
1 Introduction .................................................................................................................. 5  
2 State of the Art ............................................................................................................. 7  
   2.1 Common Semantic Framework (task 6.1) ................................................................. 8  
      2.1.1 Semantic Web Technologies - Semantic Desktop ........................................... 8  
      2.1.2 Operational Modules ................................................................................. 11  
      2.1.3 Knowledge-based resources ....................................................................... 16  
      2.1.4 Conclusion ................................................................................................. 17  
   2.2 Social and informal learning (task 6.2) ................................................................. 18  
      2.2.1 Computer Supported Collaborative Learning ............................................... 18  
      2.2.2 Social media applications and tagging in eLearning ...................................... 21  
      2.2.3 Overview of social media or knowledge sharing applications and their communities ................................................................. 26  
      2.2.4 Available technologies for data extraction from social media applications ................................................................................................. 27  
      2.2.5 Social learning platforms ............................................................................ 29  
      2.2.6 Related EU Projects ................................................................................... 29  
      2.2.7 Conclusion ................................................................................................. 29  
3 Design of services ........................................................................................................ 31  
   3.1 Integration of formal and informal learning: setup ................................................. 34  
   3.2 Task 6.1: Common semantic framework architecture (formal learning component) ................................................................. 41  
      3.2.1 Document Annotation Subsystem .................................................................. 44  
      3.2.2 Semantic Search Subsystem ......................................................................... 47  
      3.2.3 Ontology Management Subsystem ............................................................... 49  
      3.2.4 Concept/Mind Map Tool ............................................................................. 51  
      3.2.5 Use case 6.1 ............................................................................................... 53  
      3.2.6 Showcase 6.1 .............................................................................................. 54  
   3.3 Task 6.2: adding the social component ................................................................. 54  
      3.3.1 Use case 6.2 ............................................................................................... 58  
   3.4 Task 6.2a: domain knowledge & social tagging ..................................................... 59  
      3.4.1 Social Tagging ............................................................................................. 60  
      3.4.2 Relation extraction methodology ................................................................. 60  
      3.4.3 Experiment with sparse, anonymous data from Delicious ................................ 61  
      3.4.4 Applying the approach to other knowledge sharing platforms ....................... 62  
      3.4.5 Interaction between ontologies, folksonomies and learning material .............. 63  
      3.4.6 Showcase 6.2a ........................................................................................... 65  
   3.5 Task 6.2b: social networks ..................................................................................... 65  
      3.5.1 Discovery of the user's social network ............................................................ 66  
      3.5.2 Management of the user's social network ...................................................... 67  
      3.5.3 Analysis of the social network ...................................................................... 68  
      3.5.4 Search and retrieval of content from the network ........................................... 68  
      3.5.5 Design of the framework and services .......................................................... 69  
      3.5.6 Preliminary results ....................................................................................... 72  
      3.5.7 Showcase 6.2b ............................................................................................ 73  
4 Conclusion .................................................................................................................... 74
D6.1 Social and informal learning support design

5 References.................................................................................................................. 75
Appendix A.................................................................................................................... 78
Appendix B...................................................................................................................... 84
Appendix C...................................................................................................................... 89
Executive Summary

In a Lifelong Learning context, learners access and process information in an autonomous way. They might need information to be able to develop hobbies, to fill a knowledge gap or simply to keep updated with respect to new developments in their areas of interest. To this end, learners might rely on formal learning, that is they might focus on textual material approved by content providers in the context of a course developed by an organization or institution. Alternatively (or in addition), they might want to rely on informal learning, that is on (non-)textual material available through the web which is uploaded and approved by the community of learners and not necessarily by a content provider of an institution.

The goal of this Workpackage is to develop services that facilitate learners and tutors in accessing formal and informal knowledge sources in the context of a learning task. More specifically, we will support the learner through a personalized search. The system will provide recommendations on the basis of the user’s profile, his interests, his preferences, his network and obviously the learning task. An advantage of this system with respect to a standard Google search is that the list of search results will be prioritized and categorized according to the conditions specified by the learners. In addition, a trust dimension will be added since the output proposed will take into account the opinions of the user's trusted network of friends.

More specifically, we will develop a Common Semantic Framework (CSF) that will support the integration of formal and informal learning. It will allow the stakeholders to identify, retrieve and exchange the relevant learning material. Communication will be facilitated and new communities of learners will be established through the recommendations provided by the system. The CSF should support formal learning which for the time being we assume will be driven by a Learning Management System and will address mainly textual material such as textbooks, articles, slides as well as informal learning which we identify with (non-)textual material emerging from social media applications. In particular, we plan to include Wikipedia pages, videos from YouTube, images from Flickr, URLs of relevant websites from Delicious, Q&A from Yahoo answers as well as material from forums and blogs. It is through the access to formal and informal material that new knowledge will originate.

In order to achieve this goal, we rely on knowledge repositories. In the LT4eL project (Monachesi et al. 2008), we have concluded that an appropriate way to retrieve formal content might be by means of an ontology which can guide and support the learner in the learning path, facilitate (multilingual) retrieval and reuse of content as well as mediate access to various sources of knowledge. In our specific case this will be ontology in the domain of computing, which is used to annotate the learning material and to search within a Learning Management System, guiding thus the learning process.

We will also employ tags and social networks which emerge from the informal learning based on social media applications that will be used for search and recommendation. The ultimate goal is to complement the formal knowledge represented by domain ontologies with the informal knowledge emerging from social tagging (about content and users), improving thus the possibility of retrieving appropriate material and allowing learners to connect to other people who can have the function of learning mates and/or tutors. In this way, it will be possible to provide
a more personalized learning experience able to fulfill the needs of different types of learners. In order to reach these objectives, two tasks have been identified in our Workpackage:

- Task 6.1: Creation of a knowledge sharing network.
- Task 6.2: Adding a social component to the public knowledge

In this report, we describe the achievements made in the first phase of the project to reach our goals. More specifically, we report on the state of the art with respect to social and informal learning. We include a description of the necessary technology available for the realization of the Common Semantic Framework. In addition, we present an overview of the most recent literature on the use of social media in eLearning as well as the way tagging, a crucial component of social media, is being used both within these media and in eLearning. We have also included an overview of some of the most relevant social media communities and social learning platforms.

The main goal of this research was to identify the tools and methodologies that are already available and motivate our choices in adopting them for the realization of the Common Semantic Framework. In this first phase of the project we have made preliminary design choices with respect to the realization of the Common Semantic Framework and we have set the basis for how the tools and the methodologies we have decided to employ will be adapted for the development of our system which should integrate formal and informal learning. In this first phase of the project, we have worked at the formal and the informal learning components separately and we have investigates ways to connect the two. In this respect, our research has focused on the way we are going to relate informal material, tagging and communities of learners to formal learning material that has been structured by means of an ontology. Preliminary experiments have been carried out to this end and have been summarized in this report.
# 1 Introduction

In a Lifelong Learning context, learners access and process information in an autonomous way. They might need information to be able to develop hobbies, to fill a knowledge gap or simply to keep updated with respect to new developments in their areas of interest. To this end, learners might rely on formal learning, that is they might focus on textual material approved by content providers in the context of a course developed by an organization or institution. Alternatively (or in addition), they might want to rely on informal learning, that is on (non-)textual material available through the web which is uploaded and approved by the community of learners and not necessarily by a content provider of an institution. However, informal and social learning have also become part of learning situations where an organization or institution is involved given the pervasive use of social media especially among the new generation of learners.

Typically, learners employ their computer to store and organize information, they access various digital sources available on the internet and have at their disposal different communication technologies to interact with other people, depending on context, skills and preferences. In order to fulfill their learning task, they can access formal knowledge sources such as textbooks, courses, slides, articles. They can find them by means of standard search engines within a learning management system, in case they carry out their learning tasks within an institution. However, they can also access informal knowledge, that is a growing body of information available via social media such as Wikipedia or YouTube, as well as forums and blogs that might originate from the educational activities or simply from the internet. These media add a social dimension to learning since they not only provide learning material but also a community that allows for communication and information exchange about these resources, including the possibility to comment on their quality.

The goal of this Workpackage is to develop services that facilitate learners and tutors in accessing formal and informal knowledge sources in the context of a learning task. More specifically, we will support the learner through a personalized search. The system will provide recommendations on the basis of the user profile, his interests, his preferences, his network and obviously the learning task. An advantage of this system with respect to a standard Google search is that the list of search results will be prioritized and categorized according to the conditions specified by the learners. In addition, a trust dimension will be added since the output proposed will take into account the opinions of the user's trusted network of friends.

More specifically, we will develop a Common Semantic Framework (CSF) in which the stakeholders will be able to identify, retrieve and exchange the relevant learning material. Communication will be facilitated and new communities of learners will be established through the recommendations provided by the system. The CSF should support formal learning which for the time being we assume will be in the context of a Learning Management System (LMS) and will address mainly textual material as well as informal learning which we identify with (non-)textual material emerging from social media applications. In particular, we plan to include Wikipedia pages, videos from YouTube, images from Flickr, URLs of relevant websites from Delicious, Q&A from Yahoo answers as well as material from forums and blogs. It is through the access to formal and informal material that new knowledge will originate.
In order to achieve this goal, we rely on knowledge repositories. In our specific case this will be an ontology in the domain of computing, which is used to annotate the learning material and to search within a Learning Management System, guiding thus the formal learning process. We will also employ tags and social networks which emerge from the informal learning based on social media applications that will be used for search and recommendation. The ultimate goal is to complement the formal knowledge represented by domain ontologies with the informal knowledge emerging from social tagging (about content and users), improving thus the possibility of retrieving appropriate material and allowing learners to connect to other people who can have the function of learning mates and/or tutors. In this way, it will be possible to provide a more personalized learning experience able to fulfill the needs of different types of learners. In order to reach these objectives, two tasks have been identified in our Workpackage:

- Task 6.1: Creation of a knowledge sharing network
- Task 6.2: Adding a social component to the public knowledge

In this report, we will describe the achievements made in the first phase of the project to reach our goals. It is organized as follows. In section 2, we discuss the state of the art with respect to formal and informal learning. It includes a description of the necessary technology available for the realization of the Common Semantic Framework. In addition, we present an overview of the most recent literature on the use of social media in eLearning as well as the way tagging, a crucial component of social media, is being used both within these media and in eLearning. We have also included an overview of some of the most relevant social media applications and social learning platforms. The goal of this section is to identify, the tools and methodologies that are already available and motivate our choices in adopting them for the realization of the Common Semantic Framework. Section 3 describes the design choices we have made with respect to the realization of the Common Semantic Framework and discuss how the tools and the methodologies we have decided to employ will be adapted for the development of our system which should integrate formal and informal learning. In this first phase of the project, we have identified the properties of the formal learning and the informal learning components separately and we have investigates ways to connect the two. More specifically, we discuss how we are going to relate social media, tagging and communities of learners to formal learning material that has been structured by means of an ontology. Section 4 concludes the report with an overview of the achievements made in this first phase and plans for the future.
2 State of the Art

The main objective of the workpackage is to create the appropriate methodology in order to support formal and informal learning and the emergence of new knowledge. Hence, a system will be developed that should be able to identify, retrieve and recommend relevant material for a certain user on the basis of his profile, his network and the learning task. The Common Semantic Framework (CSF) constitutes the core of this supporting system.

This section is divided in two parts. In section 2.1, we provide an overview of the existing resources, tools and methodologies that will play a role in the development of the CSF. In order for the CSF to be a suitable basis for the integration of formal and informal learning, it needs to provide tools and resources to support the learners and tutors in their activities. With respect to formal learning, the CSF needs to provide at least the functionalities developed within the framework of the LT4eL project (www.lt4el.eu) and to extend them in the direction of a more convenient and personalized environment and a more precise document annotation. In order for the CSF to support informal learning, it needs to provide an extensible architecture which should allow easy transition from a private working environment to a social network environment. As a prototype for the CSF, we consider the work done on the Semantic Desktop, where the personal working space is enhanced with Semantic web technologies. In the project, we will apply these ideas to the learners/tutors working space in which they process interconnected learning materials ranging from formal learning objects, stored in an LMS, to informal recommendations or knowledge, extracted from social networks. Below we provide an overview of resources which will be built-in the CSF, such as ontologies and multilingual lexicons. They will play a role in the (multilingual) search and recommendation system. The ideas behind the Semantic Desktop is that it will allow for a more interactive communication between the web and the personal requirements of the user; annotation tools can be employed for the semantic annotation of the learning material; search engines - for the retrieval of the learning material while visualization tools are adopted for a conceptual representation of the retrieved learning material. These tools, which will be at the heart of the CSF, are investigated in order to identify the most appropriate ones to be selected for our purposes. Please note that the architecture of the CSF is described more elaborately in section 3.

In section 2.2, we focus on informal learning and summarize recent papers that describe some of the relevant components. Specifically, we have investigate social media applications, as well as tagging that can play a role in integrating formal and informal learning since tags can be used to enrich existing ontologies and finally we discuss social networks. The goal is to identify an appropriate methodology and the appropriate technology that can be adopted in the integration of the formal and informal learning components within the CSF. In addition, we have made an inventory of existing social media applications and their communities with the objective to select the ones that we will include in our system. Furthermore, we have made an overview of possible platforms that can support social learning and of existing projects that work on similar issues in order to exchange results.
2.1 Common Semantic Framework (task 6.1)

In this section, we present an overview of existent components, resources and tools that can be used for the development of the Common Semantic Framework. In defining the CSF, we are driven by the work on the Semantic Desktop, since it provides the right functionalities for the tasks which should be supported within the workpackage. It should be noted that we do not have the resources within the project to implement a full-fledged Semantic Desktop, but we will implement the functionalities which are necessary and sufficient for the tasks related to the formal and informal learning.

Thus, we first present some existing research on the Semantic Desktop. We conclude the overview with a list of services and resources that will be implemented in the project in order to construct the CSF. These services include: ontology management system; NLP resources and tools - lexicons, annotation grammars, etc; annotation tools; search engines; semantic web tools; user oriented tools like concept maps and mind maps. Some of the tools and resources will be reused from the LT4eL project (www.lt4el.eu): in particular, the ontology management system and NLP resources and tools. They will be further developed within this project. The remaining tools and resources will be developed within the project by means of adapting existing tools and resources which are reviewed below. The specific role of these tools and resources in the architecture of the system will be presented in section 3, that is Design of services.

2.1.1 Semantic Web Technologies - Semantic Desktop

Concerning next generation web technologies, the most popular ones are the Semantic Web Technologies, including tools and standards which form the basic building blocks of a system that could support the vision of a Web enhanced with meaning. The Semantic Web has been developing a layered architecture, which is often represented using a diagram first proposed by Tim Berners-Lee, with many variations since. The following figure gives a typical representation of the layered architecture.
The layered architecture allows usage (or implementation) only of some of the technologies on some of the layers depending on the application. Also, it allows the introduction of new layers necessary for the corresponding tasks.

The ideas of creating a system that can process metaknowledge in a safe and trusted way can be extended with the idea of processing multidimensional metaknowledge from distributed resources with different formats. The resources in that case will be processed in a uniform way and will be incorporated in the so-called **Social Semantic Desktop**.

The minimum set of requirements for the Semantic Desktop are as follows (Decker and Frank 2004):

1. Naming conventions
2. Standardized desktop ontologies (e.g. PIM ontologies like iCal)
3. Wrapping of legacy information
4. Multidimensional metadata and data browsing
5. Metadata storage and querying
6. Linking of data items

The Semantic Desktop is viewed as a first step towards the Social Network through P2P services. An overview of the Semantic Desktop building blocks is described in (Sauermann, Bernardi, and Denge 2005, p. 3). These blocks are: integrated projects and tools. The authors give the following definition of the Semantic Desktop:

A Semantic Desktop is a device in which an individual stores all her digital information like documents, multimedia and messages. These are interpreted as Semantic Web resources, each is identified by a Uniform Resource Identifier (URI) and all data is accessible and queryable as RDF graph. Resources from the web can be stored and authored content can be shared with others. Ontologies allow the user to express personal mental models and form the semantic glue interconnecting information and systems. Applications respect this and store, read and communicate via ontologies and Semantic Web protocols. The Semantic Desktop is an enlarged supplement to the user’s memory.

The authors also address the subjectivity of the personal conceptualization and the common ground. They consider that: … the individual background is expressed using personal mental models, expressed as personal concepts; and the common background is represented by common ontologies. Both are formalized in RDF and preferably OWL and are used by the desktop application.

The Semantic Desktop search is also outlined. Ontologies and context-based searches are viewed as central. Then, various types of user interfaces are presented.

In Sauermann (2005) the idea of transferring the web technology to view single PCs is discussed. The desktop data are represented via URIs in RDF, which is then connected to the relevant ontology. The focus is on the integration of the components.

In Sauermann et al. (2006) a concrete implementation of the Semantic Desktop idea is described, called Gnowsis. It is split in two parts: a server and a GUI. Concerning the ontology-based organization, the PIMO (Personal Information Model) is discussed. It consists of a formal part (upper, mid and domain formalized ontologies), and of a user part (extensions made by the user). The other parts – using the Web technologies for tagging, and semantic wikis – concern the informal learning component. However, the
ontology matcher component seems a good extension of the ontology towards handling the web gathered resources.

A specification of the Semantic Desktop vision is the Social Semantic Desktop (SSD) paradigm. It adopts the ideas of the Semantic Web paradigm, which offers a solution for the web. Formal ontologies capture both a shared conceptualization of desktop data and personal mental models. RDF serves as a common data representation format. Web Services - applications on the web - can describe their capabilities and interfaces in a standardized way and thus become Semantic Web Services. On the desktop, applications (or rather: their interfaces) will therefore be modeled in a similar fashion. Together, these technologies provide a means to build the semantic bridges necessary for data exchange and application integration. The Social Semantic Desktop will transform the conventional desktop into a seamless, networked working environment, by loosening the borders between individual applications and the physical workspace of different users.

The [NEPOMUK Project] aims to provide a standardized description and implementation of a Semantic Desktop architecture, independent of any particular operating system or programming language. The NEPOMUK features are:

- **Search**: enables users to search for resources amongst different and distributed sources;
- **Desktop**: On their desktop, users manage resources using integrated applications for creating or editing documents. NEPOMUK is designed to provide a notification management system for the user to receive information, regarding shared resources. Even when offline, users are able to access relevant resources transparently.
- **Profiling**: by logging the user’s activity, NEPOMUK should be trained to behave according to the specific user’s needs. This automatic behaviors must include annotations and information regarding trust with other users or sources.
- **Data Analysis**: To ease semantic annotation of unstructured documents, users can use keyword extraction, sorting and grouping of the result, according different criteria, usage of reasoning for infers new information.
- **Social**: At the social level, the management of groups and users enhances social interaction and ease resource sharing. Access rights management tackles with the security needs. Users can publish and subscribe to relevant stream of information, such as the modifications made to a particular resource or the results of a search.

We believe that the results of the NEPOMUK project might play an important role in the development of the CSF and in integrating the formal and informal components of learning. It is obvious that the functionalities of the CSF could be implemented on the basis of the corresponding services.

The following table contains a short description of the tools we have found relevant for the implementation of the Social Semantic Desktop functionalities within the CSF. Our strategy will be that if some of the tools could be easily adapted to our needs, we will use them directly. In cases where this is not possible, we will re-implement the relevant functionality within the CSF.

<table>
<thead>
<tr>
<th>Tool Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haystack</td>
<td>Haystack is a project at the Massachusetts Institute of Technology to</td>
</tr>
</tbody>
</table>
D6.1 Social and informal learning support design

<table>
<thead>
<tr>
<th>Tool</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semex</td>
<td>Semex (SEMantics EXplorer) is a system aiming to help users organize their personal information in a semantically meaningful way. Semex views personal information as a network consisting of instances, such as persons, articles, emails and pictures, and the associations between these instances.</td>
</tr>
<tr>
<td>IRIS</td>
<td>IRIS is a semantic desktop application framework that enables users to create a personal map across their office-related information objects. IRIS includes a machine-learning platform to help automate this process. It provides dashboard views, contextual navigation, and relationship-based structure across an extensible suite of office applications, including a calendar, web and file browser, e-mail client, and instant messaging client.</td>
</tr>
<tr>
<td>DeepaMehta</td>
<td>DeepaMehta is an open source semantic desktop application based on the Topic Maps standard. It aims at evolving nowadays separated desktop applications into an integrated workspace, enabling the user to organize, describe and relate information objects like text notes, external documents and media, browse the web and create semantic networks—all these in one seamless, semantic-enabled desktop environment.</td>
</tr>
</tbody>
</table>

In the next sections, we will present an overview of the tools and resources we consider as an initial prerequisite for the implementation of the Social Semantic Desktop functionalities in the CSF.

2.1.2 Operational Modules

The operational modules ensure the implementation of the support software modules: search, visualization, inference, annotation, etc. They also provide an opportunity for adding the social dimension.

This section includes an overview of:

- concept/mind map tools,
- annotation tools,
- search engines.

Concept/mind map tools are investigated in order to check whether they can be extended in order to be used within the architecture of the Semantic Desktop. We envisage the extension of such existent tools for representing various types of distributed resources in an interactive and uniform way. Our aim is to create an extension of the Semantic Web Layered Model with an additional layer that will include tools for organizing, unifying and presenting distributed resources that contain metaknowledge.
Annotation tools are assessed in order to choose suitable ones that can be used for automatic, semi-automatic and manual addition of meta or user specific knowledge to learning materials.

Search engines are reviewed in order to choose and/or extend available ones that can be used for indexing and searching in a Document Repository with annotated XML documents.

**Concept/Mind Map Tools**

A **concept map** is a diagram showing the relationships between concepts. Concepts are connected with labeled arrows, in a downward-branching hierarchical structure. The relationship between concepts is articulated in linking phrases, e.g., "gives rise to", "results in", "is required by," or "contributes to". Concept mapping is a technique for visualizing the relationships between different concepts.

A **mind map** is a diagram used to represent words, ideas, tasks, or other items linked to and arranged radially around a central key word or idea. Mind maps are used to generate, visualize, structure, and classify ideas, and as an aid in study, organization, problem solving, decision making, and writing. The elements of a given mind map are arranged intuitively according to the importance of the concepts, and are classified into groupings, branches, or areas, with the goal of representing semantic or other connections between portions of information.

Concept and mind maps are very similar in definition. If resources are linked to nodes of a mind/concept map we get a content map. A review list of concept/mind map tools is provided below. This investigation has been carried out because we aim at extending existent mind/concept mapping tools with functionalities for automatic visualization of resources like ontologies, lexicons, xml documents in a map view. These types of resources will be thus available to ordinary users who are not ontology engineers or linguists. The targeted user groups (stakeholders) are mostly tutors and learners, and such a visualization of the resources is necessary for achieving a high level of human perception.

<table>
<thead>
<tr>
<th>Tool Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>wikimindmap</td>
<td>WikiMindMap is a tool to browse easily and efficiently in Wiki content, inspired by the mindmap technique. This tool aims to support users to get a good structured and easy understandable overview of the topic one is looking for. <a href="http://www.wikimindmap.org/about.htm">http://www.wikimindmap.org/about.htm</a></td>
</tr>
<tr>
<td>freemind</td>
<td>FreeMind is a premier free mind-mapping software written in Java. The tool supports: browsing the local file system as a map, creation of a mind map from scratch, addition of local and remote hyperlinks to a node, merging of maps, insertion of encrypted nodes, browsing of mindmaps, configuration of the tool, nodes formatting, searching for a node by label. It is integrated with a web browser. <a href="http://freemind.sourceforge.net/wiki/index.php/Main_Page">http://freemind.sourceforge.net/wiki/index.php/Main_Page</a></td>
</tr>
<tr>
<td>VUE (Visual Understanding Environment)</td>
<td>VUE is a 2D drawing application specialized for concept/ content maps creation and presentation. It supports the following features: linking of resources to concept map's nodes; browsing of linked to</td>
</tr>
</tbody>
</table>
nodes resources; search for resources/nodes by specified keywords, categories, labels; multiple resources linking to a node; creation of presentation pathways - linear and non linear; playback a created pathway presentation; map display tools: zooming tool, full screen view, outline view, split screen view of multiple maps, pruning; keywords and categories definition for nodes; search on keywords and categories in conjunctive and disjunctive mode; ontologies loading, attachment of ontological concepts to nodes; merging maps in voting or weight fashion.

http://vue.tufts.edu/index.cfm

**SmartDraw**

It has the following features': Assisted Flowcharting; Creation of Business Graphics; PowerPoint Integration with Auto Animation; Graphic Design; Charting; Image Charts; Interactive Maps.

http://www.smartdraw.com/

**Thinkgraph**

ThinkGraph is a mix between a 2D Drawing application and a Concept Map editor. The tool supports: 2D Drawing(creation of 2D basic shapes: Line, Rectangle, Ellipse, Polygon, Text, Bitmap Image (PNG or JPEG)); formatting of text shapes: edition of font size, font color and style (Bold, Italic, Underlined); rotation of shapes; bookmarking of web resources.


**VYM (View Your Mind)**

VYM (View Your Mind) is a tool to generate and manipulate concept/ mind maps. The tool supports the following functions: 2D Drawing - creation of 2D basic shapes: Line, Rectangle, Ellipse, Polygon, Text, Bitmap Image (PNG or JPEG); formatting of text shapes: edition of Font Size, Font Color and Style (Bold, Italic, Underlined); addition of icons, bookmarks to nodes.

http://www.insilmaril.de/vym/

**Kdissert**

Semantik (previously Kdissert) is a mindmapping-like tool to help users to create: presentations, dissertations, theses, reports in an interactive and understandable view. The tool is available for Windows, Linux and other free operating systems.

http://freehackers.org/~tnagy/kdissert.html

On the basis of the investigation we have carried out on the properties of the various tools, we have concluded that we will adapt the **VUE** and **freemind** concept/mindmap tools for our purposes.

**Annotation Tools**

An annotation tool is software for manual or automatic annotation (tagging) of documents. Annotation is add-on information asserted with a particular point in a document or other piece of information. We envisage that some languages automatic annotation of documents will be available. The users will be supplied with an interactive manual annotation system where more specific and user dependent
annotations might be added to the documents. Some available annotation tools are reviewed below.

<table>
<thead>
<tr>
<th>Tool Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SHOE Annotation Editor</strong></td>
<td>The Knowledge Annotator is a Java program that allows users to mark-up web pages with SHOE knowledge. The Annotator is available as an applet or a stand-alone Java application. Documents are annotated with SHOE Ontologies (<a href="http://www.cs.umd.edu/projects/plus/SHOE/spec1.0.html">http://www.cs.umd.edu/projects/plus/SHOE/spec1.0.html</a>). The annotation includes creation of instances and relations. <a href="http://www.cs.umd.edu/projects/plus/SHOE/KnowledgeAnnotator.html">http://www.cs.umd.edu/projects/plus/SHOE/KnowledgeAnnotator.html</a></td>
</tr>
<tr>
<td><strong>Annotea protocol</strong></td>
<td>Annotea is a framework that supplies a Java library for communicating with the Annotea server. Annotations, according to Annotea, mean comments, notes, explanations, or other types of external remarks that can be attached to Web document. When the user gets the document he can also load the annotations attached to it from a selected annotation server or several servers and see what his peer group thinks. <a href="http://www.w3.org/2001/Annotea/">http://www.w3.org/2001/Annotea/</a></td>
</tr>
<tr>
<td><strong>Annotea client implementation: Annozilla</strong></td>
<td>Annozilla is designed to view and create annotations associated with a web page, as defined by the W3C Annotea project. The idea is to store annotations as RDF on a server, using XPointer-like constructs. The intention of Annozilla is to use Mozilla's native facilities to manipulate annotation data. The functionality, that is supplied, is: loading of annotation; creation of annotations; posting annotation to an annotation server. <a href="http://annozilla.mozdev.org/index.html">http://annozilla.mozdev.org/index.html</a></td>
</tr>
<tr>
<td><strong>Annotea client implementation: Amaya</strong></td>
<td>Amaya is a Web editor. Browsing features are integrated with the editing and remote access features in the environment. Amaya supports annotation of HTML, XML, XHTML, MathML, and SVG documents. The application is based on Resource Description Framework (RDF), XLink, and XPointer. <a href="http://www.w3.org/Amaya/">http://www.w3.org/Amaya/</a></td>
</tr>
<tr>
<td><strong>Annotea client implementation: Annotea Ubimarks in Mozilla/Firefox</strong></td>
<td>Annotea Ubimarks is part of Annotea social bookmarks and topics work in Mozilla. It makes users familiar with the common bookmark user interface metaphora to create metadata for Semantic Web, to share and combine bookmark data and bookmark categories, or topics from several locations or with other metadata. Topics in Annotea can be simple tags or they can form hierarchies. <a href="http://www.annotea.org/mozilla/ubi.html">http://www.annotea.org/mozilla/ubi.html</a></td>
</tr>
</tbody>
</table>
| **Annotea server implementation: Zope Annotation Server** | This product provides an annotation server for Zope, implementing the W3C's Annotea protocol which allows clients to add annotations to web pages without modifying the pages' source. Zope is an open source web application server. It features a transactional object
D6.1 Social and informal learning support design

database which can store not only content and custom data, but also
dynamic HTML templates, scripts, a search engine, and relational
database (RDBMS) connections and code.
http://www.zope.org/

Annotea server
implementation:
W3C Annotea
Server
Annotea is a framework that supplies a Java library for communicating
with the Annotea server. Annotation clients need to follow the Annotea
protocols and understand the annotation schema. The user can install
his own annotation server. The Annotea server is a general purpose
RDF store, with a general Algae query interface and a couple of
optimized queries for annotations, threads etc.
http://www.w3.org/1999/02/26-modules/User/Annotations-HOWTO

CLaRK
CLaRK is an XML-based software system for corpora development,
maintenance and manipulation. It is implemented in Java. It
incorporates several technologies: XML technology; Unicode; Regular
Cascaded Grammars; Constraints over XML Documents.
http://www.bultreebank.org/clark/

For the task of the automatic documents annotation, we consider CLaRK as the most
appropriate tool, because it was successfully used within the LT4eL project. If
necessary, we will consider using other tools as well.

More details concerning the listed tools can be found in Appendix A.

Search engines
A search engine is an information retrieval system designed to help find information
stored in a documents repository. Search engines provide an interface to a group of
items that enables users to specify criteria (search query) about an item of interest and
have the engine find the matching items. There are several styles of search query
syntax that vary in strictness. Some search engines apply improvements to search
queries to increase the likelihood of providing a quality set of items through a process
known as query expansion. We envisage the following requirements for a search
engine to be used in the LTfLL project:

- supports query expansion
- supports indexing of XML documents and searching within such created indices
- supports searching in different parts of the XML documents
- supports retrieval of different parts of matching XML documents

We review some possible candidates below:

<table>
<thead>
<tr>
<th>Tool Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oracle Berkeley DB XML</td>
<td>Oracle Berkeley DB XML is an open source, embeddable XML database with XQuery-based access to documents stored in containers and indexed based on their content. Oracle Berkeley DB XML is built on top of Oracle Berkeley DB and inherits its rich features and attributes. Oracle Berkeley DB XML adds a document parser, XML indexer and XQuery engine on top of Oracle Berkeley DB. <a href="http://www.oracle.com/database/berkeley-db/xml/index.html">http://www.oracle.com/database/berkeley-db/xml/index.html</a></td>
</tr>
</tbody>
</table>
### Lucene Based Search Engine

The Search Engine is written in Java and uses the Lucene API (integrated in Google) for documents indexing and retrieval. The Engine indexes XML documents by using a user defined indexing schema (also an XML document) which contains identifiers (lucene fields' IDs) and pointers (XPath pointers which point to a content of the XML document that will be the Lucene fields value). The fields have a hierarchy structure. The searched fields can be of any depth, the retrievable ones are the leaves. According to the schema definition the Engine can be used for semantic or text-based search. The documents can be retrieved by evaluating queries which are XML documents defined according to a specific format. The Search Engine is integrated in the CLaRK System and it was used as a component in two other projects: LT4el and AsIsKnown.

We are developing a Search Engine based on Lucene which should operate on an XML repository and our goal is to integrate it in the CSF.

#### 2.1.3 Knowledge-based resources

The Knowledge-based resources are presented here briefly, for the sake of completeness, since these are mostly re-used from a previous related project - LT4eL. Needless to say, these resources will be enriched and adjusted for the purposes of the present project. The adjustment might not be trivial in all cases. It should be stressed that the ontology is assumed to be the driving force behind the Common Semantic Framework. The Semantic based technologies and the Operational modules are discussed in more detail, because our objective is to evaluate which tools would be the most appropriate for the development of the CSF.

The following knowledge-based resources have been considered:

- ontologies
- language lexicons: domain-specific as well as more general ones
- learning objects + metadata

Concerning the knowledge-based resources, we will rely on the outcome of the EU project LT4eL (www.lt4el.eu). The aim of this project was to enhance eLearning with Language and Semantic web technology in order to develop innovative applications for education and training. It focused on the integration of the semantic knowledge in eLearning, i.e. the use of ontologies, a key element in the Semantic Web architecture, to structure and retrieve the learning material within the LMS. The learning objects were annotated on the basis of the ontology which can be used also for cross-lingual retrieval and for ontology-based search.

An explicit connection was established between: a top ontology and a domain ontology; the ontology, language-specific lexicons and learning objects; the ontology, grammars for annotation of concepts, and learning objects. It also gives insights on the interrelations among pure text tokens, keywords, definitions and concepts.

Within the project a domain ontology in the domain of 'information technology for end users' was developed. The domain ontology has been linked to the DOLCE upper ontology and to OntoWordNet. In addition, the ontology has been linked to domain specific lexicons which have been developed for eight of the languages of the consortium and learning objects have been annotated on the basis of these concepts.
Semantic search has been implemented on the basis of this setup, as well as ontology browsing and cross-lingual search.

At the moment, the ontology contains 1002 domain concepts, 169 concepts from OntoWordNet and 105 concepts from DOLCE Ultralite. It also contains more than 100 object properties. Using the model of ontology-to-text relations, concept annotation grammars were developed for Bulgarian, Czech, Dutch, English, German, Polish, Romanian. This model reflects the relation between concepts and text by means of the annotation grammars. In the model, it was assumed that the ontology represents the lexical meaning of the language. Thus, for each concept the lexical items and non-lexical phrases that represent the content of the concept were detected.

There were two problems: (1) for some of the concepts in the ontology, there exists no lexical item in one of the target languages; (2) for some of the selected lexical items, there was no concept in the ontology that represents the right meaning. The first problem was overcome by allowing also non-lexical (fully compositional) phrases to be represented in the lexicon. The second problem was solved by the extension of the ontology.

The ontology-to-text model became the prerequisite for the creation of the semantic search module. We consider this status of developed resources a good basis for the further extensions within the LTfLL project. These extensions will provide facilities for improved semantic annotation. This improved semantic annotation will provide the basis for a more efficient semantic search and implementation of recommendation system that should facilitate the learning process.

### 2.1.4 Conclusion

We have provided an overview of the existent resources and tools that can be integrated and/or adapted for the development of the Common Semantic Framework. This Framework is meant to support learners in their life-long learning process. Our assumptions are:

1. to re-use some already developed tools from a previous related project (LT4eL),
2. to rely on contemporary techniques to achieve a learner-friendly educational environment.

With respect to the re-usage, the following resources and tools are envisaged:

(a) knowledge-based language resources, such as ontologies, language specific lexicons and annotation grammars.

(b) NLP tools, such as keywords and definitions extractor, linguistic processing chains, semantic search engine.

All these will be integrated into the CSF. They will be also extended, adapted, re-structured to support additional functionalities within the new task.

With respect to current techniques which might play a role in the development of the CSF, the following technologies and approaches are considered suitable:

(a) Semantic Web Technology, which is behind the idea of the Semantic Desktop. The implementation of the CSF will use a Semantic Desktop approach. This would ensure the learner's personalization of the web applications and information. It can be implemented either as a stand-alone application, or connected to an LMS. The concept/mind map tool(s) will be the main mechanism for organizing different types of semantic resources. Filtering will be added, which is based on the user profiles. The
learners will be given the facility to translate the material chosen by the ontologies. They will have the possibility to annotate the learning material, and to explore the whole user workspace of resources and social connections. For that purpose, the following tools have been selected: VUE and FreeMind.

(b) The investigation of the existing annotation tools has shown that there aren't any annotation tool that can satisfy our basic requirement: sound annotation of XML documents with knowledge-based and linguistic information which should be domain independent and should be easily configured for different domains and languages. Concerning automatic and semi-automatic annotation, the CLaRK system is viewed as a candidate. Its functionality can be used for performing basic tasks associated with the CSF, such as linguistic processing chains, XML documents transformation, grammar annotation. The CLaRK system is developed by the IPP-BAS partner team, so extensions are envisaged according to the LTfLL requirements. Where available appropriate external tools will be integrated for specific processing.

(c) We also focused on the available information retrieval technologies and existent applications. We have been developing a search engine application based on Lucene API, which is considered as an efficient text search implementation. Alternative solutions like Oracle Berkeley DB XML are also taken into account. Our aim is to incorporate in CSF an efficient search engine application that can be integrated into other systems and can also operate on XML documents repository, irrespectively of the XML documents validation schema.

Our conclusion is that there is a great variety of knowledge-based and language resources as well as operational modules that have been used for tasks similar to ours, but most of them do not satisfy the basic requirements for CSF. Thus, the suitable tools have to be extended in order to be integrated successfully into the CSF system.

2.2 Social and informal learning (task 6.2)

In this section, we provide an overview of the issues that we have investigated in order to develop a methodology and an architecture for the integration of social and informal learning in the Common Semantic Framework. More specifically, we have reviewed the literature on social learning with the aim of establishing the theoretical foundations of our task, that is which theory will be at the basis of our work. Furthermore, we have investigated the literature on social media applications (and tagging) to assess to which extent they are already employed in eLearning. We have also carried out an analysis of the most relevant features of these applications in order to decide which ones should be included in the project given their content, their tags and the networks of people using them. In addition, we have made an overview of commonly used web and social media technologies that can be employed in our implementation. We have also identified other EU project that could be relevant for our task with whom we could exchange results.

2.2.1 Computer Supported Collaborative Learning

This section introduces the most important theories in the field of Computer Supported Collaborative Learning (CSCL). These theories represent the pedagogical and psychological foundation for the major developments in this field. The improvements that we are going to bring to the domain of social learning should have their basis on the major theories in CSCL. We describe the role of the society and cultural environment in a person's education, the concept of community of practice
D6.1 Social and informal learning support design

and the way this supports learning, as well as the ways in which knowledge building takes place inside these communities.

Offering support to the learner in a social environment is a branch of computer supported collaborative learning (CSCL). The most relevant ideas from this research area are presented below and they constitute the basis of our research.

Vygotsky’s socio-cultural theory (Vygotsky, 1986) contains the most influential ideas in this domain. The basic assumptions behind his theory are that the most important factors in someone’s development are the social and cultural factors. A person constantly learns from the other people in his society. Not only is he taught how to behave during childhood, he also forms his ideas based on his social environment and develops them using his group’s support. The culture of the larger group of people to which an individual belongs is another important factor in his development.

From Vygotsky’s ideas many other theories evolved within CSCL. One of them is the theory of Situated Cognition (Lave & Wenger, 1991). This idea is based on the concept of “community of practice” which is actually a social group just like in Vygotsky’s theory.

The theory of situated cognition says that the individual learns by being part of such a community of practice. Starting as a novice situated on the periphery of the community, the individual evolves to an intermediate user who has a more central role in the community; finally, the intermediate user becomes an expert, whose role is the most central in the community. The process of the learner’s evolution in the community represents the learning process by which the learner becomes knowledgeable about the community’s artifacts. The concept of artifact represents an item, concept or even metaphor that has a specific particular meaning for the members of a community and that helps the members of the community to accomplish their activities. The concept of artifact is important for Vygotsky but also for some of the exponents of the “theory of activity” like (Engeström, 1987). Another important concept from CSCL is the one of “Knowledge building communities” (Scardamalia and Bereiter, 1994).

Another type of community is the one identified by Berlanga et al. (2008). They define the ad hoc transient community as "a community which exist to fulfill a particular request (their ad hoc-ness) and for a limited period of time only (their transience)." An example of such a community would be providing support upon request to a group of learners. An important characteristic of such a community would be that "knowledge sharing is not imposed, it arises spontaneously, even though technology may be used to guide and speed up the emergence of the community". The authors also identify three necessary conditions for knowledge sharing to occur:

- The community should have a clear goal.
- The community should have members with different levels of knowledge in the domain of the community - experts, intermediates, novices.
- The community should monitor the members' participation and performance - the accountability criteria.

The main idea of the CSCL theory of learning in communities of practice is that a class of students should perform like a community and the activities of the students should be mainly to build knowledge. An important idea is also that the dialog between teacher and students should be typical to a community and not the classical “teacher asks, student answers, teacher evaluates” which is typical to a classical
learning environment. Another important theory in CSCL is Problem Based Learning (PBL). This argues that learning happens as students solve problems that appear in their path.

We consider very important also the vision of Stahl (2006) on the knowledge-building process. He views this process as a complex interaction between the personal understanding and the social knowledge building, as presented in the previous image. Stahl also describes the main CSCL tools (forms, wikis, glossaries) that can be used together in order to implement this cycle.

From Stahl’s “Group Cognition: Computer Support for Building Collaborative Knowledge” we can adopt the following ideas regarding the implementation of CSCL tools:

- The knowledge building process occurs in small groups of people 3-6 persons.
- The implementation of a learning solution in a real environment is often “sabotaged” by the users’ desire to use their familiar tools like personal e-mail, personal and not integrated instant messaging, other editing and publishing tools, other communities.

To conclude, we consider the following ideas valuable for the LTfLL project:

- Both children and adults develop knowledge in a social environment;
- Learners can develop knowledge by becoming members of a community of practice, by getting to know the community's artifacts and by evolving to a more central role inside the community
- Students build knowledge together in the classroom acting as a community and the dialogue between student and teacher changes accordingly
- A personal idea is brought to the social level where it is discussed and it may become an artifact of the community
• The Situated Cognition theory of Lave & Wenger, (1991) fits nicely with the social media applications and network setup which will be at the basis of informal learning and can be thus adopted for our task.

2.2.2 Social media applications and tagging in eLearning
A growing number of articles have been published on social media applications and the use of tagging, which are two aspects that are going to play an important role in social learning and in the integration of informal learning in the Common Semantic Framework. However, our conclusion is that their use in (lifelong) learning has not been fully exploited yet with most applications still at the experimental level. We provide an overview of the state of the art in this area by focusing on the work which can be relevant for our tasks. More specifically, we discuss literature about the social dimension in eLearning and we summarize the relevant work on social tagging, social media and their use in eLearning pointing out why we consider it important for the development of the CSF.

Marenzi et al. (2008) suggest that social media will change the traditional way in which systems for learning and collaboration work, since people use social software individually, and not within the framework of a clearly defined group or project. They describe three example scenarios, and then discuss different kinds of ties between people and applications that support each situation:
• Strong ties / group collaboration: tools like BSCW and Wikis are appropriate
• Weak ties (people who are acquainted but don't have to work together) / information propagation: classic email and social networking platforms such as Facebook
• Potential ties: discussion forums, blogs with search functionalities, Q&A sites
• Collective intelligence: use collaborative tagging with a very large number of people

Finally, they discuss Facebook as a possible social networking platform in which all the functionalities needed for learning can be integrated, as Facebook allows the integration of other applications. This is an insight which might be also relevant for the LTfLL project and which will be discussed more in detail in section 3. More specifically, we believe that the CSF should operate on existing social media applications and social networking platforms and we will investigate through questionnaires whether Facebook might constitute a good basis to create our subcommunity of learners.

Similar relations are discovered also by Angehrn and Maxwell (2008) during the development of TENTube - A video-based connection tool supporting competence development. The authors claim based on existing research that a person can only have genuine social relationships with 150 people. TENTube was designed to motivate users to establish connections that do not exist by creating awareness, stimulating interest, and providing a pretext for making new connections. From this tool we can identify the following relations between users, videos and tags:
• Video is related to Competence/Tag
• User has submitted/seen Video
• Video has inspired Video
• User knows User
These relations can be used to further identify the relations between peers identified by Marenzi and presented above.

An example of existing social media that falls under the category of informal learning is discussed by Bernhard and Gurevych (2008). They propose a way to find answers to users' questions, making use of social Questions & Answers websites such as WikiAnswers, Yahoo! Answers and AnswerBag. The available data consists of pairs of a posed question and the given answer. Before publishing a user's question, such systems propose to view the answer to a similar question. WikiAnswers also provides a way for a user to indicate that such a similar question is actually what he meant, in which case his question will be saved/annotated as a paraphrase of the original question. An experiment is described where similarity measures (string-based and vector-space-based) are applied to questions from WikiAnswers and question paraphrases of the original question, while the objective is to find the original question for each of the question paraphrases. The idea is to use those measures on new paraphrases to find an existing question. The answer can be shown unchanged. An accuracy around 80% and mean reciprocal rank around 90% are reported for the vector-space-based measures. Linguistic pre-processing (lemmatization, spelling correction etc.) does not generally improve the results. We plan to employ Q&A content in our project as part of the informal learning and we believe that their methodology to find similar answers can also be useful in our project.

Artail et al. (2008) experimented with a collaborative learning environment in which a knowledge base of questions and answers is built up by means of user contributions. A credit/debit system is used to make sure that users not only pose questions but also answer other users' questions. A rating system for answers is coupled with the credit system: providers of answers get more credits as their contributions are rated better; providers of a rating get more credits as the rating they gave is closer to the average rating for the regarding contribution. Furthermore, the system uses NLP (mainly sentence similarity) to find similar questions and answers. The principles of rating and credit/debit points fit well our objectives: rating helps the system to recommend materials and helps the user to decide, while a credit/debit system can be used as a motivation for the learners to contribute to knowledge. We plan thus to implement a similar system as one of our services in the CSF.

Grosseck (2007) focuses on the social bookmarking system del.icio.us and its use in eLearning. The article gives an overview of possibilities, advantages and disadvantages of del.icio.us. Four styles of tagging are distinguished (quoted from a non-English website article by Javier Cañadas): selfish, friendly, altruist, popular. Subsequently, some ways are identified of using del.icio.us to support learning and teaching. Basically, the mentioned points boil down to the practice that a teacher as well as his students use del.icio.us to save and tag bookmarks, possibly always using one tag that is agreed on to identify the course, in addition to tags describing the bookmarked resource. A teacher can recommend materials by tagging them with the course identifier; students can subscribe to an RSS feed for such a tag. Students can collaboratively collect additional materials, and can recommend resources to each other by actively posting a saved/tagged bookmark to an acquainted user. Teachers can monitor what the students are saving and how they tag. The collection of stored links related to teachers and courses will be preserved also if the course ends and will grow with each following course, so that future teachers and students can profit from
it. We plan to exploit del.icio.us for its content and its tags within the CSF and we have already carried out experiments to this extent.

Bateman et al. (2007) also argue that collaborative tagging fits well with eLearning, and give some more precise advantages:

1. it supports self organization of learning content – motivation for learners to do tagging
2. potential to further enrich peer interactions and peer awareness
3. tagging gives learners the opportunity to reflect on content and summarize it
4. the tags provide insight on the learner’s comprehension and activity – tags represent expertise

They also report on an experiment where three sets of tags for learning materials are compared: tags provided by students, tags provided by teachers, and the result of an automatic metadata extractor. Some of the findings are, that half of the teachers’ tags were not given by students, while 68% of the automatic metadata extractor tags were among the students' tags. Furthermore, the authors introduce the annotation and tagging system OATS, which allows for highlighting besides tagging. The result is collaborative highlighting, so users can see which parts of a text are highlighted by most peers. This study is particularly relevant for our project since it shows that there is a big difference between tags provided by experts and those provided by non experts and both are necessary to be able to locate resources. It supports our insight that domain ontologies which drive the learning process in our setup should be enriched with social tagging in order to be useful both beginners and advanced learners.

In this first phase of the LTfLL project, we have exploited tagging mainly to access and extract knowledge from social media and to establish a link between users, concepts and resources. However, we envisage exploring the use of tagging also for more educational purposes.

Vuorikari (2007) investigates the use of tags in a multilingual setting. She describes an experiment where a group of teachers, independent of each other, are asked to bookmark and tag material that they are collecting. There were teachers of various languages. 26% of the given tags were in English, even though none of the users were native English speakers. Furthermore, 15% of all tags were helpful in all the languages: this concerns tags that have roughly the same spelling in many languages, such as names. Also, the tags were categorized into factual, subjective and personal, where the category "factual" is subdivided into "topic" and "category refinement". 79% of all tags were classified as topic tags, 14% as category refinement (thus 93% factual); the remaining 7% were subjective tags. Since in the LTfLL project, we are dealing with a multilingual situation, this article has given us some insight on the use of multilingual tagging.

Social tagging is not just an alternative way to organize resources. It also reveals emerging collective knowledge, such as which words are good keywords for a specific resource, which topics are popular among a community and which words are related to each other. The underlying classification structure is often referred to as a folksonomy (Vanderwal 2005). Mika (2005) and Lambiotte and Ausloos (2005) describe folksonomies as a tri-partite model, where each ternary edge represents the fact that an actor (user) relates a tag (concept) to a resource.
Social tagging systems vary with respect to a number of properties that all have an impact on the content and usefulness of the resulting folksonomy. Marlow et al. (2006) have created a taxonomy of tagging systems: they define a number of key dimensions that characterize tagging systems. Those dimensions are:

- Tagging rights (who can add and remove tags? E.g. resource owner, tag creator, everyone)
- Tagging support (are any tag suggestions given while users provide new tags?)
- Aggregation (bag or set of tags: are unique tags given by different users counted?)
- Type of object (e.g. web pages, images, videos)
- Source of material (e.g. uploaded by user, any web resource, or system proposed what should be tagged)
- Resource connectivity (are resources linked to each other or grouped?)
- Social connectivity (which kind of explicit connections are there between users, and are they symmetrical?)

An important distinction which was already identified by Vanderwal (2005) and relates to the dimension of tagging rights, is broad/narrow folksonomy. In broad folksonomies, many people tag the same items, which gives additional collaborative information about a specific item. In narrow folksonomies, the owner of an item is the only one who tags it.

A number of publications remark that social tagging is not only a flexible way of classification because no pre-defined vocabulary is needed (e.g. Marlow et al. 2006), but in fact a way to discover the shared vocabulary of a community (Marenzi et al. 2008), which can include many new community-specific terms that are not yet included in existing lexical resources (Cattuto et al. 2008). Sen et al. (2006) describe how three different aspects influence the evolution of tag vocabulary in a system: personal tendency, community influence and the tag selection algorithm for tag suggestion. Personal tendency is most likely to cause diversity in the vocabulary, while community influence contributes to convergence. One of the conclusions is that "community influence on a user’s first tag is stronger for users who have seen more tags". The strength of community influence is affected by the tag selection algorithm and is lower if fewer shared tags are suggested. Marlow et al. (2006) show a related but different effect: there is a bigger overlap between the vocabularies of users who are connect through their contact lists then between two randomly chosen non-related users.

Commonly cited is the article by Golder and Huberman (2005) about the structure of collaborative tagging systems. They discuss the difference between tagging and traditional classification with a taxonomy, and describe a number of problems in this respect introduced by tagging, i.e. polysemy, synonymy and basic level variation (the basic level means how general or specific a concept makes most sense to a person, e.g. should it be “bird” or “robin”, but this varies among persons). Earlier tags in a bookmark tend to represent basic levels. The authors also provide an overview of several functions tags have for bookmarks:

- tags that identify the topic of the bookmark - these are the most relevant tags as they refer to the content of the document
- tags that identify the type of content (article, book, video, blog)
tags that identify the owner of the content - these are mostly irrelevant for search but can be useful for identifying groups inside the community.

• tags that have no individual meaning but help to refine another tag/category

• tags that identify characteristics of the content. They are mostly adjectives and usually provide a personal positive or negative vote for the specific content.

• tags that identify the relation to the tagger (usually “my…”)

• tags that are used by the tagger for self organizing like "todo" for example.

Furthermore, it was determined, based on a data set from delicious, that after around 100 users bookmarked an item, the relative frequencies of the tags stabilize. The insights of the reviewed papers on social tagging have played an important role in determining the design of the CSF and the way we plan to integrate formal and informal learning, as described in more detail in section 3.

Some publications discuss relations between tags. Specia and Motta (2007) argue that groups of related tags can be used for query extension, or as suggestions to users during tagging. Their goal is to make explicit the semantic structure of a folksonomy for semantic web applications, and describe an approach that uses tag preprocessing (morphologic similarity, exclusion of isolated tags), statistical tag clustering based on co-occurrence, and relation identification by looking up terms in online ontologies. Van Damme et al. (2007) describe a similar approach, which however focuses on how an actual ontology can be generated on the basis of a folksonomy. They propose that in addition to providing the tags, the community can also directly help to identify or judge/approve relations in the ontology. This is an approach we also plan to adopt in the LTfLL project. Both approaches make use of online lexical resources: in order to discover whether terms are acronyms, misspellings or variations, both use Google and Wikipedia, while Van Damme et al. use the Leo Dictionary and Wordnet in addition.

Schmitz (2006) and Heymann and Garcia-Molina (2006) developed algorithms to derive a hierarchy of tags, based on the data of a tripartite tagging network rather than on existing lexical or ontological resources. Cattuto et al. (2008) compare three measures for relatedness that are also applied directly to the folksonomy structure: one measure based on co-occurrence, one based on the cosine similarity, and the FolkRank algorithm (Hotho et al. 2006). The measures are applied to find a set closely related tags. Subsequently, the authors propose a mechanism of semantic grounding: the found tags are mapped to WordNet, in order to inspect the semantic distance between the related tags. Cosine similarity appears to yield more synonyms, where the other two measures rather yield different concepts, among which are super concepts, which make them appropriate for retrieving taxonomic relationships. FolkRank, in addition, is capable of detecting multi-word lexemes from distinct tags.

The topic of multi-word lexemes as tags has not been given much attention in the literature. We observed that in some systems, tags have to be separated by commas in the input field, some systems take into account a multi-word tag if it is put between quotes, and some consider every space a delimiter and do not allow multi-word tags. The latter can have consequences for the popularity of (parts of) multi-word tags, as people deal with them in different ways. E.g. the tag “web development” is written “web-development” or “webdevelopment”, so that both of them have a smaller number of users than would be the case with a unique spelling. To make it worse, some users type the space anyway, thus contributing to the use of the individual tags “web” and “development”, so that the tags “web-development” or “webdevelopment”
receive an even lower count. These effects will be taken into account by the service of the CSF that analyzes existing tagging data. The service to tag learning materials within the CSF will allow for tags that include a space.

### 2.2.3 Overview of social media or knowledge sharing applications and their communities

We have investigated the most relevant social media applications available and the communities which have grown around them in order to assess which role they could play in the Common Semantic Framework. The criteria for selection were:

- Users share a specific kind of information with the whole community, e.g. bookmarks, or media such as videos. So we exclude websites like www.facebook.com that are mainly focused on a social network for friends.
- Communities mentioned in scientific literature about folksonomies and social tagging are included.
- Small communities are not included if bigger community exists from which the same kind of information can be retrieved.

We selected the following platforms for further investigation: Delicious, Digg, Diigo, Shelfari, YouTube, blogs (no specific platform), Twitter, Flickr, Yahoo Answers.

We investigated for each system the most important aspects, according to the following categorization of social tagging systems by Marlow et al. (2006):

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Main categories by Marlow et al. (2006)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tagging Rights</td>
<td>• self-tagging</td>
</tr>
<tr>
<td></td>
<td>• permission-based</td>
</tr>
<tr>
<td></td>
<td>• free-for-all</td>
</tr>
<tr>
<td>Tagging Support</td>
<td>• blind tagging</td>
</tr>
<tr>
<td></td>
<td>• suggested tagging</td>
</tr>
<tr>
<td></td>
<td>• viewable tagging</td>
</tr>
<tr>
<td>Aggregation Model</td>
<td>• bag of tags</td>
</tr>
<tr>
<td></td>
<td>• set of tags</td>
</tr>
<tr>
<td>Object Type</td>
<td>• textual</td>
</tr>
<tr>
<td></td>
<td>• non-textual</td>
</tr>
<tr>
<td>Source of Material</td>
<td>• user-contributed</td>
</tr>
<tr>
<td></td>
<td>• system</td>
</tr>
<tr>
<td></td>
<td>• global</td>
</tr>
<tr>
<td>Resource Connectivity</td>
<td>• links</td>
</tr>
<tr>
<td></td>
<td>• groups</td>
</tr>
<tr>
<td></td>
<td>• none</td>
</tr>
<tr>
<td>Social Connectivity</td>
<td>• links</td>
</tr>
<tr>
<td></td>
<td>• groups</td>
</tr>
<tr>
<td></td>
<td>• none</td>
</tr>
</tbody>
</table>

We replaced "Social Connectivity" by "Explicit Social Connectivity", i.e. relations that users actively establish between each other, such as joining a group or adding someone as a friend. This is distinguished from another category that we added,
"Derived Relations", which contains relations that are derived by the system, such as a community of users who have all used a certain tag.

The complete overview of the selected systems can be found in Appendix B. We have decided to include Delicious, YouTube, Flickr and Yahoo Answers to extract relevant content as well as tags and user networks (when available) for integration in the CSF, since their properties fit our needs, as discussed in more detail in section 3.

The emergence of a huge number of social networking and social media sites gives rise to systems that combine the social sites, for users who are active on many of them. They collect the information the user is monitoring and notifications about new events (e.g. when the user receives a message through one of the social sites, or a blog he follows is updated). Examples of such platforms are http://8hands.com/ and http://www.ubervu.com/, the latter being in preparation for its first launch.

We will apply a similar methodology in the architecture of the CSF, where information from different social media applications are combined, however we focus on the retrievable content and relations between users rather than new events.

2.2.4 Available technologies for data extraction from social media applications

Various technologies have emerged that provide access and allow easy development of services for the social networking applications. This section is going to present the most relevant of them, technologies that we envisage to use for the integration of the informal learning component in the Common Semantic Framework.

The social networking applications offer APIs that allow collection of data or interaction with the application. Those APIs are based on XML or on XML-based technologies.

For data extraction purposes those applications usually offer web feeds. Web feeds are a format (usually XML-based) for providing data for users or third party applications. The most common formats for distributing web feeds are RSS and Atom. RSS originally came from Rich Site Syndication and it is now the most used format for distributing short summaries of the updated content for rich content web sites. RSS is still used in many versions across the internet - 0.9, 1.0, 2.0 and RSS 1.0 is based on RDF one of the most useful languages for the semantic web.

Atom is a newer standard for web feeds that offers support for multilingualism and unicode. Atom is used by many content providers including for example the feeds offered by Google.

Besides these standards employed for distributing feeds, relevant information can be extracted from metadata that is represented through the following formats:

- FOAF - Friend of a Friend is an RDF based standard for describing personal profiles and connections with other persons. FOAF is used for example by one of the largest social networking communities - Hi5.
- XFN - XHTML Friends Network is a much simpler standard than FOAF. XFN allows description of the relations between 2 persons using just the rel attribute of an HTML Anchor element. Through the values of this element, a user can specify if the person to which the link is pointing is a friend, was met, is a colleague and many kinds of relations. The XFN standard is used mostly inside blogs and its values are used for example by Google Open Social.
D6.1 Social and informal learning support design

(http://code.google.com/apis/opensocial/), a Google platform for discovering social connections.

• SIOC - Semantically-Interlinked Online Communities (http://sioc-project.org/) - is an ontology for describing the social web data in RDF. The SIOC ontology is a W3C Member Submission and is on its way to become a standard. As it is still a very young technology it has not been yet adopted by the large content providers but it can be used to export and to import to/from different applications.

In order to extract or interact with the data from the social networking applications we might need to use an authentication protocol. The most used protocol for authentication in the social web APIs is OAuth (http://oauth.net/). According to its authors, OAuth is "an open protocol to allow secure API authorization in a simple and standard method from desktop and web applications." Furthermore, the authors report that it is used or in development in the following companies: "Digg, Jaiku, Flickr, Magnolia, Plaxo, Pownce, Twitter, and hopefully Google, Yahoo, and others soon to follow."

Other resources that might be useful for our task are:

• Freebase - http://www.freebase.com - a knowledge base containing data about concepts in the world around us and even specific concepts from the computer science and medical field. Freebase provides an API for querying the database using their own specific query language MQL.

• DBPedia (http://dbpedia.org/) is a RDF version of Wikipedia that can be queried using SPARQL. DBPedia contains most of the knowledge in Wikipedia expressed using a semantic web technology.

• Sindice (http://sindice.com) is an index of semantic data gathered from the internet. It is claimed that "over 10 billion pieces of reusable information can already be found across 100 million web pages which embed RDF and Microformats". Sindice offers an API based on OpenSearch that provides results using RDF, Atom or JSON.

Another kind of technologies that can be used inside our project are the ones that help users connect on different web sites not depending on a specific technology or platform. This idea that originally started a while ago with Microsoft Passport is now best represented by OpenID (http://openid.net). OpenID is a technology that allows a user to have a single identity across the internet and only use that identity for many websites. The difference between OpenID and Microsoft’s Passport is that OpenID is distributed across more content providers and a user can choose his most trusted provider for hosting his data.

Using technologies like OpenID, Google Friend Connect (http://www.google.com/friendconnect/) and Facebook Connect (http://wiki.developers.facebook.com/index.php/Facebook_Connect) allow users to use data from Google and Facebook in their own private web site in order to connect to their friends and to bring them to other sites of interest. Facebook Connect and Google Friend Connect are for the moment just released or in limited preview release but they suggest the direction where these companies are heading from the social networking point of view. Concerning the directions that need to be followed in the field of data exchange between applications, the dataportability (http://dataportability.org/) project that aims to develop best practices and to highlight
the best technologies that allow people and applications to share data on the social web.

2.2.5 Social learning platforms

In order to support social and informal learning and combine it with formal learning, we need to choose an appropriate platform that would give us the necessary functionality. To this end, we have carried out an investigation of existing platform that could be used in our project. There exist many systems that offer a virtual learning environment. An overview of learning management systems can e.g. be found on wikipedia: [http://en.wikipedia.org/wiki/Learning_management_systems](http://en.wikipedia.org/wiki/Learning_management_systems).

However, such systems are mostly used within the context of an educational institution, and cannot be exploited and used by everyone. On the other hand, open platforms where people actually can learn something appear to be mainly forums, and do not offer support for management of learning material or activities, other than basic communication and file sharing features that some forums have.

A special kind of forum where people answer others’ questions is a Question&Answers (Q&A) website. In contrast to a traditional forum, where people ask for advice, give and share opinions and experiences, the purpose of a Q&A site is to provide a specific answer to a specific question. Examples of such sites are [http://www.answerbag.com/](http://www.answerbag.com/), [http://wiki.answers.com/](http://wiki.answers.com/) and [http://answers.yahoo.com/](http://answers.yahoo.com/).

A smaller, somewhat similar site with community is [http://www.questler.com](http://www.questler.com), an "informal learning network". Forum topics are called "quests" and an interesting feature is that the principle of social tagging is applied: quests are tagged, whereas in the aforementioned Q&A websites, questions are put in a category hierarchy.

Furthermore, we did not come across social network sites that focus on learning material and learning as their main activity. There are eLearning communities or social networks dedicated to eLearning on e.g. [http://eduspaces.net/](http://eduspaces.net/), [http://classic.elgg.org/](http://classic.elgg.org/) and [http://moodle.org/](http://moodle.org/), but the communities seem to be about lifelong learning at a metalevel, discussing about learning and conceptual, functional and technical issues related to the eLearning software.

2.2.6 Related EU Projects

Several projects are emerging at the European level that can produce interesting results which can help us reach the objectives of our Workpackage. We have an ongoing collaboration with the following projects in order to exchange results and information:

- APOSDELE - learn@work
- iCamp - the educational web for higher education in an enlarged europe
- MELT - Metadata Ecology for Learning and Teaching
- BONy - Babylon & ONtology
- TenCompetence - Building The European Network for Lifelong Competence Development

More information about these EU projects can be found in Appendix C.

2.2.7 Conclusion

In this section, we have investigated some relevant issues for the integration of informal learning in the CSF. We have identified important references in the field of
Computer Supported Collaborative Learning (CSCL). More specifically we have concluded that the Situated Cognition theory of Lave & Wenger (1991) constitutes a relevant theoretical basis for the way informal learning could take place in our system. We expect that content emerging from social media applications and the communities emerging from social networks are going to play a relevant role in informal learning. This assumption is in accordance with the idea defended by Lave & Wenger that the individual learns by being part of a community of practice and that his role evolves from beginner to expert.

Furthermore, the overview of the literature on social media and tagging that we have carried out, reveals that there is an obvious potential in using social media applications for learning purposes. However, at this stage, discussions are mainly carried out in the research community and only prototype systems of certain applications seem to be available. The potential of social tagging as a medium for learning and for categorizing knowledge is also addressed in the literature and we envisage a relevant role for it in linking the formal and informal learning process. The formal learning process will be driven by the ontology while the informal one will be driven by the tags. By merging tags and ontologies, we will be able to support search and recommendation both for the beginners and the advanced learners.

We have also established that among actual users and existing systems, there is a distinction between eLearning systems, websites that host social networks, and knowledge sharing websites. Systems where a certain kind of resources are shared, e.g. Flickr or YouTube, tend to use social tagging, while knowledge in the form of Q&A or forums is not tagged in most cases. Systems that focus on learning are often used in formal learning settings, using local installations related to an institution. Information for informal learning, on the other hand, is accessible by everyone, but it is not very organized, yet since it can be available from any forum, blog or ordinary website. A more organized form of informal learning can be found in Q&A websites, however they lack the advantages of tagging. The CSF would thus fill this gap by providing a way to merge the formal and informal learning processes by providing more structure to the informal learning while opening up formal learning to the various web communities and resources.

We have carried out an investigation to assess which social media applications would be suitable for our task and we came to the conclusion that Delicious, YouTube, Flickr and Yahoo Answers will be included, in addition to Wikipedia which has a special status being a resource that creates a bridge between formal and informal content. We have also made an inventory of the techniques available to access social media applications and networks in order to extract relevant content and users (and the relation between the two). They will become relevant in the second phase of the project when we will be working at the integration of informal learning in the CSF.

Even though the social web is growing at a fast speed and technologies to exploit social media applications are evolving, there is still no prominent initiative that is going to exploit the content and network of these applications for learning purposes making our project and our tasks still very innovative.
3 Design of services

The amount of material that can be used for learning purposes is growing and it might go beyond formal educational content such as textbooks, articles, exercises or presentations originally developed by educational institutions. Due to the available technology, learners can also access more informal educational content which is accessible through social media applications such as Wikipedia, Yahoo Answers, YouTube, Flickr, Delicious, as well as forums and blogs that might originate from the educational experience. These repositories feature a strong social component and allow users to share and reuse the uploaded material as well as to comment on its quality.

In the LT4eL project (Monachesi et al. 2008), we have concluded that an appropriate way to retrieve formal content might be by means of an ontology which can guide and support the learner in the learning path, facilitate (multilingual) retrieval and reuse of content as well as mediate access to various sources of knowledge. Domain ontologies might be a useful support in a learning path since they provide a formalization of the knowledge of a domain approved by an expert. However, this formalization might be too static, incomplete or might not correspond to the representation of the domain knowledge available to the learner which might be more easily expressed by the tagging emerging from communities of peers via available social media applications.

It would be desirable for a learner to have access to a system that could combine formal and informal knowledge sources in order to be able to fulfill the needs of different users. It is thus necessary to access different repositories of knowledge which might be local to a Learning Management System or available on the Web. This will be highly heterogeneous data which can be manipulated through a resource ontology. In addition, we need to create a link between the formal representation of a given domain in the form of ontologies and the informal descriptions produced by social tagging. Furthermore, we will create a community of learners by giving suggestions of peers that might fulfill the role of a tutor or of learning partners for a given learning task.

The main objective of the workpackage is to create the appropriate methodology to integrate formal and informal learning. We build an infrastructure for knowledge sharing in which learners can develop a system of interoperable personal and community knowledge-bases which is best formalized by means of an ontological layer. More specifically, we are going to develop services that are based on the interaction between a formal representation of domain knowledge and a social component which complements it. The traditional bipartite model of ontologies will be extended leading to a tripartite model of users (actors), tags (concepts), and resources (instances of concepts) (Mika 2005), including thus the social dimension into this system.

In order to reach these objectives, two tasks have been identified in this workpackage:

- Task 6.1: Creation of a knowledge sharing network
- Task 6.2: Adding a social component to the public knowledge

Task 6.1 is mainly concerned with the creation of services to support social and informal learning by developing an appropriate infrastructure. The goal is to allow the retrieval and sharing of the acquired knowledge between learners and tutors. This will go beyond a simple collection of articles, textbooks, presentations, exercises, etc and
it should include all the various stages of the knowledge discovery which should be the final goal of any learning activity.

We are going to specify and implement a common semantic framework (CSF) that allows for identification, retrieval, exchange and recommendation of relevant material as well as for the relevant discourse which will allow communication among users. The CSF will be ontology driven allowing thus for a formalization of the knowledge arising from the various stages of the learning life-cycle. This will include also a formalization of the common knowledge of the domain, in a way that can support sharing and collaboration, as well as personal and community knowledge-base construction. In addition, knowledge of the resources involved as well as of the pedagogical context should be encoded and formalized through an ontology.

An annotation tool will be employed to allow learning material to be annotated on the basis of the ontologies (content, pedagogical, resources) which constitute the backbone of the common semantic framework. In this initial phase, we assume that the learning material including the ontologies and annotation should be shared among the learners belonging to the community which are granted access to it. Experiments will be carried out to assess whether the annotation process can be indeed carried out by students and whether it could be part of the learning process. We could envisage several degrees of complexity in the annotation task, from simple tagging (cf. task 6.2 below) which should be feasible for everybody to ontology driven annotation which would allow for a deep annotation of the learning objects on the basis of the relevant concepts present in the ontology. In addition, we aim at offering the possibility to choose between manual, semi-automatic and automatic annotation.

More specifically, the system should allow for the coexistence of the private knowledge of the single learner and the public knowledge which might emerge from the interaction within a given community of learners. Socialization plays an important role in this knowledge infrastructure since it can lead to the emergence of a public view on a subject based on the private views made available to the community. In addition, a learner may combine what he knows with the public view on a certain topic to discover something new. A side effect of this annotation activity is that the learning material will be associated to the appropriate metadata so that it would be possible to retrieve it successfully and to recommend it to new learners.

Summarizing, the main objectives of task 6.1 will be:

- to implement a common semantic framework which allows communication among users as well as identification, retrieval and recommendation of relevant material;
- to develop an annotation tool that will allow learning material to be annotated
- to annotate learning material (i.e. textual and visual) on the basis of the ontologies which constitute the backbone of the common semantic framework
- to set the basis for a knowledge sharing network which can be used to support formal and informal learning and the emergence of new common knowledge

Task 6.2 is concerned with providing a social component to knowledge. Social tagging systems are gaining popularity among web users which annotate their resources (web pages, images, videos, etc.) with a set of tags/keywords, which they believe will be relevant to characterize the resource according to their own needs and eventually retrieve them. This allows for an evolving classification structure which is often referred to as folksonomy (Gruber 2005, Shirky 2005), that can also play an important role within eLearning systems. The idea behind folksonomies is that users
are allowed to describe a set of shared objects with a set of keywords of their own choice, in our case, the objects will be learning material, which the learners can tag with respect to content but especially with respect to their learning experience (i.e. whether suitable, interesting, boring, too difficult, etc.). Networks of folksonomies can be represented by a tripartite model which includes the set of actors (users), the set of concepts (tags, keywords) and the set of objects annotated (learning material). It is through tagging that learners with similar interests and preferences can be identified. Tags/keywords can be viewed as the element which connect resources to learners, in this way it is possible to identify relationships among users (several users may use the same tags) and among resources (resources can be tagged with the same words).

The main goal of this task is to merge the dynamic knowledge provided by users/learners through tagging with the formal knowledge provided by the domain ontologies by adding tags/concepts (or instances of concepts) and relationships (or instances of relationships) between concepts in the domain ontology. Furthermore, actors will be included to extend the bipartite model of ontologies which includes concepts and instances (Mika 2005). We assume that collection of tags can emerge both from the learners activities within the system but they can also be extracted from those social media applications that provide them such as Delicious or YouTube. To this end, we apply standard techniques to extract the popular tags and the related tags which are included in the ontology. In this way, we create a link between the resource (instance of concept), what it is about (concept) and who has provided it (actor).

In order to develop a proper knowledge sharing framework, we need to be able to create and populate domain ontologies semi-automatically on the basis of the available material. To this end, the existing concepts in the ontology will trigger the identification of new related concepts from the tags available from the social media. Relations are identified by employing existing databases which have relations among concepts stored by users. In addition, Natural Language Processing (NLP) techniques will be employed to extract possible relations from corpora on the basis of the available tags. Given the new possibilities that have emerged in the area of knowledge extraction from social media we have decided that it would be more fruitful to adopt this approach to carry out ontology enrichment with new concepts instead of an NLP approach as originally envisaged in the Description of Work.

However, it is not always the case that tags are provided to identify resources in social media. In the case of Q&A sites or in the case of blogs and wikis, tags are not available. It is in this context that NLP techniques can prove very useful. They can be adopted to extract terms/concepts, definitions and relations from the relevant material. We build on results obtained within the LT4eL project in which a keyword extractor has been developed mainly to facilitate the semi-automatic generation of metadata, while definition extraction has been employed mainly for the creation of lexica (Monachesi et al., 2006). In the context of this task, however, the keywords extracted could be used as a first step towards the creation of concepts for a domain ontology related to the learning objects adopted (cf. also recent work of Sclano and Velardi (2007)). Similarly, definitions extracted from the texts employed could be used to define the concepts present in the ontology (cf. also Klavans and Muresan (2001) and Fahmi and Bouma (2006)) and to identify relations (cf. Schutz and Buitelaar (2005)). Tags/keywords identified in this way will lack the link with the community of users.
Another way to support the social aspect of learning is to employ social media applications and the communities which are part of them. An additional objective of task 6.2 is to identify the communities that the user is part of, to help him manage them and to help him evolve inside the community, by creating and improving useful connections. We assume that the learner will not only be part of the community established within our system but could also be a member of other communities on the web.

As already mentioned, in this setup we establish an obvious link between the network of users, tagging and content. In fact, we envisage the possibility to add the social component to search by ranking on the basis of user's appreciation which might be available through tagging.

More generally, the recommendations the system will provide to the learners can be viewed as an appropriate categorization of the search results. This will be carried out on the basis of ontological information (about domain, resource and learning context), that is the ontology provides the formalization necessary to structure the heterogeneous data. The results will be prioritized on the basis of the available information coming from the social network and the user profile.

Summarizing, the main objectives of task 6.2 will be:

- to add the learner dimension to formal knowledge by integrating social tagging;
- to support social and informal learning by creating a link between the formal representation of a given domain in the form of ontologies and the informal descriptions produced by social tagging and folksonomies.
- to employ NLP techniques to extract domain knowledge from learning material for which tagging is not available (i.e. to extract concepts, definitions and relations), as well as to identify appropriate relations among tags extracted from social sites.
- to support social learning by including social networks which emerge from social media application and within our system.
- to recommend relevant peers and relevant material related to the learning task into consideration.

### 3.1 Integration of formal and informal learning: setup

In this first phase of the project, we have worked towards the design of the architecture of the CSF and have set the basis for its various components as can be seen in the picture below:
At the same time, we have worked on two crucial aspects to provide support for social and informal learning. They are tagging and social networks. More specifically, we tested a methodology for finding related tags from tagging data, which can be applied to extract knowledge from existing large social media applications as well as to tagging data that will be created through the tagging facilities that will be part of the CSF. We have also explored ways to associate this extracted knowledge to the existing domain ontology on computing. Furthermore, we have developed a methodology to identify the communities within social media applications that the learner is part of, found ways to help him manage them and to help him evolve inside the community, by creating and improving useful connections. All these aspects are spelled out in more detail in the next sections of this report.

As already mentioned, the Common Semantic Framework should be able to deal not only with formal learning objects, as is the case at the moment (i.e. textual and visual) but also with informal learning material and networks which originate from the social media applications. In addition, it should be possible to integrate the knowledge coming from tagging to knowledge which is already present in the ontology. Besides tagging, there must be support for annotation of learning material using the ontology. Furthermore, the CSF should be able to deal with the social networks within the learning community as well as from the communities emerging from external social media applications. The combination of all this available information with the specific learning task and the specific user profile will lead to a personalized search, recommendations and a visualization of content.

The functionality developed in this project will be available in the form of services that deal with input and output in a way that will be agreed upon with the other
Workpackages in the project, so that they can be integrated in one architecture. The functionality offered by our Workpackage will thus be composed of such services. The role of the Common Semantic Framework is to host and integrate the various components that provide the services, while their interface will be designed according to the guidelines and agreements that will be realized by the Integration Workpackage (WP2). The picture below provides a more detailed overview of our architecture which includes also the various services provided:

Figure 3. Relation between Common Semantic Framework, Learning Management System and data from the Internet. The key data and services and the direction of data flows are shown.

The basic resources that will be used from Internet as input for CSF are Learning Objects (LO) and information from social media applications – tags, users, tagged resources. The CSF will get as input Learning Objects, user profiles and learning tasks from an LMS.

The CSF services are of three types:

1. Automatic services: LOs content, tags, social networks analysis
2. User Knowledge contribution: LO classification, LO tagging, rating, LO annotation, Ontology feedback, Document Annotation
3. Synchronous support services that are called by the user interaction services; they do not need a user interface: users’ history control, points service

Services that will support implementation, integration and communication of the presented services are:

- Service for addition of a LO to the internal for the CSF repository
- Ontology management service
- Service for giving recommendations
- Semantic Search service
Visualization and navigation service

The resources that will flow through the CSF are stored in two repositories:

- Ontology repository: contains filtered version of the ontology for separate users
- Main Repository: contains LOs, users’ data, social networks data, tagging data

Some of the services will be developed within task 6.1 and some of them within task 6.2. More information about the services can be found in the respective sections about task 6.1 and task 6.2 (sections 3.2 and 3.3).

The interaction with the user will occur through widgets on an LMS. The widgets must communicate with the user interaction services. In the end, the services must be suited to enhance a given learning situation and infrastructure. In current practice, Learning Management Systems play an increasingly important role. However, especially since we are dealing with social and informal learning, also existing social media sites can play a role in the setup. We have identified the following possibilities for the relation between communities, social media applications and/or an LMS:

1. **Existing social media applications sites**
   The CSF operates on existing social media applications.
   
   **Main features**
   - Community: We need to create a sub-community of an existing social media application community, for example we could create a sub-community within Facebook.
   - Learning material: Consider the tagged items to be learning material. E.g. social bookmarking – bookmarked websites would be learning materials.
   - Language: Most of the existing material and tags is in English. New material and tags in sub-community can be any language.

2. **Existing LMS: new installation + new community**
   The CSF operates on a new installation of existing LMS software; a new community has to be formed that will use this installation.
   
   **Main features**
   - Community: Create an ad-hoc community.
   - Material: It has to be uploaded by initiators and community.
   - Language: We can choose one or more languages.

3. **Existing LMS + existing community**
   The CSF operates on an LMS that is actively used and which has its own community. The LMS should be enhanced with the appropriate social software.
   
   **Main features**
   - Community: Use existing community of learners
   - Material: Learning materials will be already available.
   - Language: Any language possible, usually there is one community using one language per LMS installation.

4. **Social media application that focuses on learning + existing community that tags learning materials**
   - We have not found such a community.

At the moment, we believe that the best setup which would be also compatible with the rest of the project is to combine aspects of option 2 and option 1. We expect that
an LMS will be used for the whole project which will be enriched with services
developed in WP4, WP5 and WP6. This LMS will be populated with learning objects
emerging from the project as well as with learning objects developed within the
LT4eL project. In addition, there will be a community using it for the purpose of
validation. This setup is similar to that in option 2.

However, as part of the informal learning component that will be integrated in the
CSF, the existing social networks of a learner will be taken into account, as described
in more detail in the rest of this report. In addition, relevant learning material
available through the various social media applications will be taken into
consideration. In other words, people and learning material external to our LMS will
be related to the learners that use our services. In this way, external resources such as
videos, images, bookmarks, Wikipedia pages, answers to questions, blogs and forums
are considered learning material, as people use them in informal learning. This
corresponds to option 1, with the difference that it is not necessary for our entire
community to be a sub-community in one specific existing social network. Also
information about popularity and ratings can play a role. We might revise this
decision, if it appears that our community of learners is in fact a sub-community of an
existing social network, such as Facebook. We will thus monitor carefully the
development of social media sites and investigate the status of our validation learners
through questionnaires.

We believe it is crucial to establish the link with the communities and the material
outside the LMS which might be locally used. Only in this way, we provide a genuine
social dimension to learning where collaborative knowledge from internet users is
supported. In lifelong learning, people access and process information in an
autonomous way. They access various digital sources available on the internet and
have at their disposal various communication technologies to interact with other
people. They do not want to be limited to the LMS even though the LMS might
provide an appropriate framework to give shape to knowledge acquired in a learning
dimension.

Conclusion:

5. Existing LMS, enriched with a link to material and people coming from social
media applications

Main features

• Community: Learners within LMS, but they can profit from knowledge by
  communities outside of LMS.
• Material: Partially learning objects local to our system, partially links to materials
  emerging from social media applications.
• Language: Mainly English but other languages can be included within the LMS
  local community.

In the picture below, we present a different perspective of the architecture of the CSF
and the communication with the Front End LMS. It represents a raw component
diagram in a preliminary form. Decisions are not final especially with respect to the
communication protocols and interfaces, but we have made an attempt to identify the
interconnections between the different components of the CSF and LMS.
Figure 4

The core component is the Common Semantic Framework. It integrates other components, among which are supporting components that will not have a direct connection to the Front End which we envisage as a web based system - probably a Learning Management System - to be chosen, according to the requirements specified for the LTfLL Project. It will be the central component to which other components will be integrated by means of widgets. It will communicate with external components via predefined communication protocols like SOAP, RSS or Atom.

The components that will be integrated into the CSF and/or the Front End are described below.

- **Document Annotation Component**: it will support automatic annotation of LOs implemented via language processing chains.
- **Semantic Search Engine**: the engine will get as an input the annotated XML multimedia documents and index the content of interest in them. It will supply query creation and evaluation functionality, combined with information extraction and retrieval.
- **Ontology Management System**: this component will support processing, translation, filtering, visualization and inference in ontologies. It will be used by the Semantic Search and Document Annotation components in order to perform semantic annotation and search activities over semantically annotated documents. They will access it through the Front end and/or an API. The Ontology Feedback component will communicate with the Ontology Management Component and supply the result feedback to the Front End to present the result to the users.
- **CSF DB Access Assistant**: this component serves as a bridge for the data coming from social tagging, social networks and rating systems in order to be stored in the CSF database and to be available for search, rating and browsing. It will be integrated in the Front end and will supply a common interface or separate interfaces for the mentioned components in order to upload data to the CSF.
- **Document Preprocessor**: this component will be reused or extend an existent one employed in the LT4eL project. Its function will be to convert different formats of documents to a unified XML format.
- **Image Annotation**: this component is a standalone tool for a manual annotation of images from multimedia documents with semantic information from a domain ontology, rendered by a lexicon in different languages. The tool will be extended
in order to communicate with the CSF via RSS/Atom and/or web services – SOAP.

- **Document post processor**: this component is optional. It will be used, if necessary, to clean documents after an annotation before saving them in the LO CSF Repository on the LMS side.

- **Concept map visualizer**: this component is a standalone system that will be used for visualizing concept mapping views of different types of resources: LOs, lexicons, ontologies, user workspaces (including connection to users / user groups and resources repositories). A concept mapping editor will be supplied that will be used for manual document annotation. The system will communicate with the Front End LMS via RSS/Atom protocols. The editing of a concept map view of resources will be accomplished on the user's desktop where the application is installed. The user will be offered an option to export an HTML read-only representation of the concept maps that will be stored on the LMS server and can be browsed by a logged user from any other desktop, no matter whether the standalone application is installed or not. The concept map visualization tool will be integrated with a search functionality supplied by the Semantic Search Engine component. Ontology management operations will be supplied for inferring, filtering over concept maps of ontologies.

The following components use the database of the CSF; the interaction is supported by the CSF DB Access Assistant.

- **Users History Assistant**: this component will contain history data for all the users’ interactions with other users and resources and will assist the extraction of reports for the users’ activity, rating.

- **Tag Analyzer**: this component will analyze tags available to the system in order to enrich the ontology with them and will search the web for new tags.

- **Social Networks Analyzer**: this component will analyze users’ data from social networks accessible through the LMS and new ones.

- **LO Tagger**: this component will be used for tagging LOs

- **LO Rating System**: this component will support rating of LOs on the basis of social tagging, concept annotation, user ratings

It should be noticed that in this initial phase of the project, there has been an overlap between the Common Semantic Framework and the development of the formal learning component which has been carried out under task 6.1 to the point that the two can be identified. Such overlap was also present in the original Description of Work in which we envisaged the CSF as an extension of the semantic search component proposed in the LT4eL project (www.lt4el.eu). The LT4eL project dealt only with textual documents uploaded in an LMS and approved by a content provider, thus with formal learning. In this first phase of the project, the work on the CSF which has been carried out under task 6.1 has been mainly concerned with the formal learning component. It is obvious that certain services envisaged in the formal learning component should also belong to the CSF (such as the semantic search or the visualization of the resources) which in fact can be viewed as an extension of the formal learning component.

In the next phase, we plan to develop the CSF further and integrate the informal learning component according to the architecture we have sketched in this section. The informal learning component has been developed in parallel through the activities
carried out in task 6.2. More specifically, in the next phase of the project, we envisage a division of the activities of this Workpackage in three tasks:

- Task 6.1: Creation of a knowledge sharing network which integrates formal and informal learning (UU, IPP-BAS, PUB-NCIT)
- Task 6.2: Development of an informal learning component based on social media applications and social networks (UU, PUB-NCIT)
- Task 6.3: Development of a formal learning component based on textual material contained in an LMS (IPP-BAS).

In the rest of this report, we provide a more detailed description of the activities carried out in the context of task 6.1, which has focused on the formal learning component and those services that are going to be the core of the CSF and 6.2, which has focused on the informal learning component.

### 3.2 Task 6.1: Common semantic framework architecture (formal learning component)

As mentioned in section 2 – State of the Art, the Common Semantic Framework (CSF) will provide support for both formal and informal learning processes:

- The formal processes are the ones that are guided by a tutor and where the learners are supported by tutors and advisers within the educational activities. In this kind of processes an essential role is reserved for the tutor who chooses, organizes and annotates materials via searching, concept annotations, metadata and social tagging attachment. Then he also verifies the results.
- The informal process lacks the tutor as a leader. In this process the learner manipulates the system in an attempt to find the best solution of a problem via searching within social tagging annotations, concept annotations and metadata. The learner can share his opinion and comments on a learning material by annotating it with tags and concepts and making it accessible to the community.

Irrespectively of the type of the learning process, each eLearning environment faces the need for a common semantic framework, within which the stakeholders (tutors, learners) can communicate efficiently and with the investment of a minimal effort. However, the reality is far from such an elaborated framework, concerning the connection established between the formal and informal way of learning. Practically, they interleave very often, so we envisage the task to build a framework, which relates in a better way users-to-users and learning material-to-users. It also provides a good integration of the formal education modules with the informal ones. In achieving these goals, we rely on the ideas behind the Social Semantic Desktop providing a common framework for the annotation of documents, accessing them and sharing them. The common framework will use knowledge repositories (e.g. ontologies) and the appropriate chains of NLP tools and resources to create an architecture for the knowledge acquisition and the presentation purposes.

In the available eLearning Systems, we have identified the following problems:

- lack of coordination and facilitation of the communication among users (with common interests and problems);
- lack of a common resource manipulation model that supports users’ interactions with distributed resources;
- inefficient extraction of relevant resource segments that match best to users’ search criteria.
These problems have to be solved in order to develop a better learning environment to be achieved with respect to both types of learning - formal and informal. Additionally, we have identified the following requirements to the Common Semantic Framework. The CSF has to:

1. support domain specific knowledge models aligned with appropriate lexicons (multilingual interactions);
2. support filtering of a specific knowledge domain (cultural and social adaptation of the knowledge domain to different users) in order to present a user oriented view of common knowledge models;
3. support semantic annotation, discourse analysis, disambiguation, indexing, metadata creation, LO similarity distance calculation tools (core services that will supply qualified LO creation and manipulation) in order to make the processing of semi-automatic resources possible;
4. support a unified resource model to which the different resources can be mapped (extensibility of the learning environment with new types and formats of resources);
5. support an efficient search engine that can be adapted to different kinds of search within different formats and types of resources;
6. support a common visualization of resources, which is interactive and easy for perception;
7. support a knowledge model for the user's network. This model will allow extensions for universities and their user's interaction policies;
8. provide a number of tools that will process the resources content and make user’s access to resources possible;
9. provide a number of tools for user’s communications and interactions;
10. support conceptual independent models that will define a common conceptualization for the different scientific domains (domain ontologies);
11. support a knowledge model for the resources network. This model will facilitate the adaptation of new resource formats and types as well as their manipulation by the available tools.

In order to meet the requirements listed above, we design the architecture of the Common Semantic Framework in a layered way. The layers correspond to the different elements of the technology behind the Social Semantic Desktop. At the first, very basic level, it comprises a network of resources and tools. Resources include ontologies, lexicons, learning materials, communication notes, comments, web links, etc. Each resource can have internal and external elements. Internal elements will be represented in an XML format within the CSF. The external elements will be stored outside of the CSF. The external elements will be pointed by some of the internal elements of the resource to which they are attached. The pointing to an external element will be performed by an appropriate URL. Usually, the resources within the CSF will be named by their common names, such as documents, notes, links, web pages, depending on their type. However, in the CSF they will have a similar representation. Tools are categorized as resource specific and general. Resource specific tools are those that operate over documents with a specific XML schema and manipulate particular types of documents. For example: the inference engine processes ontologies, the lexicon management tools translate ontology items via lexicons. The external elements of a resource can be accessed and manipulated via corresponding external tools, such as viewers, browsers, editors, etc. General tools are
D6.1 Social and informal learning support design

various systems for XML processing, including XML editors, transformation tools, etc.

Additionally, the resources within the CSF are interlinked. Links can be established between elements of resources, and might be either named, or anonymous. In the above sense, the links can be considered resources. The various kinds of links may be interpreted by the corresponding tools in various ways.

The main user interface to the CSF is a Concept/Mind map tool which can be customized to the different kinds of resources. Using this tool, the learner can navigate through and manipulate the space of interrelated resources and tools. The user will be able to establish new links among resources, create new resources or delete them. When necessary, the resource content can be edited with the corresponding tool.

At this layer, the CSF will be equipped with an XML oriented search engine. It will support searches on the base of the resource structure and among resources of different kinds. The search engine works on an XML repository in which all the resources are stored. The repository itself could be a private place for a user, or shared by several users. The shared repositories are the way in which the communication of resources is done. The external parts of the resources might allow additional ways of communication which are supported by the corresponding tools.

In some cases the content of a given internal element may depend on the content of some external element to the resource. In this case, the CSF will provide appropriate tools that ensure interleaving of content between these elements (when it is possible). For example, if an LO in MS Word format is updated, then its content is extracted and stored in an appropriate internal element. If an internal element is connected to a web page and its content is modified, then some update of the web page might be done. These tools will be resource dependent.

In order to ensure platform independence, the programming language of the developed and adopted tools is assumed to be Java.

In order to add the common semantic layer to the CSF, we have to extend the basic XML resources and tools layer with specific resources and tools which to cover (as a minimum) the following subsystems (cf. fig. 4):

- Document Annotation Subsystem
- Semantic Search Subsystem
- Ontology Management Subsystem
- Concept/Mind Map Visualization Subsystem tuned to the specific resources.

Additionally, there are three types of repositories: Document Repository, Index Repository, System Repository. In the Documents Repository the annotated XML documents (e.g. learning materials) are located. In the Index Repository the indices produced by the Semantic Search Engine are located. In the System Repository resources used by the CSF: ontologies, lexicons, system configuration files are located.

These resources and tools will provide the facilities for the formal learning process. Also, they are the basis for the next layer of the CSF which will support the informal dimension of the learning process.

In the rest of this section we present the main elements of this layer of the CSF.
3.2.1 Document Annotation Subsystem

Description

The Document Annotation Subsystem provides services for semantic annotation of multimedia documents. Usually, these multimedia documents will be learning objects, but in general they could be other kind of documents. These documents will be the resource for searching in by learners. The annotation of a document requires several processing steps. First, the original document can be in several formats. The most popular ones are PDF, RTF, HTML. The original document in most cases will be an external element to the CSF resource. Hence, the original document will be first converted into one or more internal elements that are parts of the XML representation of the resource. The internal elements will represent the textual content of the original documents and links to the images included in the original document. Note that images will be stored as external elements of the resource.

The subsystem provides two annotation services: "text annotation service" and "image annotation service".

Text Annotation Service

The text annotation service provides linguistic processing chains for some languages. Minimally, a linguistic processing chain (called also a language chain) will include: tokenization, POS tagging, lemmatization, semantic annotation. We will start with the language processing and semantic annotation, since these steps have already been implemented and tested within the LT4eL project. In order to improve the annotation and then the search, based on it, we will extend the linguistic annotation with discourse analysis. It would include discourse segmentation and specification of the relations between the segments. Our main goals are: (1) investigating the possibility
D6.1 Social and informal learning support design

for refining the concept recognition and sense disambiguation of the targeted words (lexical terms) via coherence relations (discourse relations, rhetorical relations) markup, and (2), enriching the granularity of learning materials to which the user would have access. The creation of the discourse annotation component will be done in several steps that may be iterated. These are as follows:

- **Creation of a coherence relations taxonomy.** We will start with the set of relations and the coding scheme defined by Wolf & Gibson 2006. Their taxonomy is based on the Hobbs list (Hobbs 1985), and is more coarse-grained than others that include up to 400 types of relations, which makes it really applicable for real time manual annotation. It consists of ten types of coherence relations: temporal sequence, cause-effect, condition, elaboration, example, similarity, contrast, generalization, violated expectation, same-segment. Same-segment is a structural type of relation, because it holds between disconnected parts of one discourse segment (subject NP separated from its predicate). Same-segment, similarity and contrast relations are symmetrical while the rest are asymmetrical (directed), that is – one of the segments is more important (the nucleus) than the other (satellite).

  According to the coding scheme the three general steps of the annotation process are: (1) the output of the sentence-splitter is segmented further into clauses and then, if needed, annotators insert intrasentential boundaries for smaller discourse segments; (2) the discourse segments are grouped thematically and (3) the coherence relations between the segments are indicated. After the annotation process is finished, the taxonomy may be further adjusted to improve the descriptive adequacy for the texts in the Computer Science domain.

- **Manual annotation of the learning objects.** We first will annotate part of the learning material by hand in order to study the impact of the discourse annotation to the solution of the above goals. Also, the manual annotated materials will be used in the developing of the discourse analysis module.

- **The analysis of the obtained discourse structures will provide information that could be used for (1) the development of constraints over the semantic annotation grammar, (2) improving the definition finder and (3) supporting anaphora resolution. In addition, we consider the possibility to create a rule-based grammar for recognizing the coherence relations that are unambiguously linguistically marked. According to Schauer (Schauer 2000) 15 to 20 percent of coherence relations are signaled by some kind of conjunction but not all of them are unambiguous.

In order to improve the concept annotation, we will test different knowledge-based techniques that are common for word sense disambiguation. Our main goal is the enrichment of the concept annotation grammar in order to map the relations between text chunks, recognized as carriers of the concepts, with relations present in the domain ontology: is-a, part-of, etc. In the future, different algorithms for automatic establishment of lexical chains (with nouns) may be tested (see Mihalcea 2007 for an overview). Lexical chains and rhetorical relations, the two types of discourse information, contributing to the text coherence, will be used for improving the concept annotation for learning materials and, subsequently, definition extraction. For example, a discourse segment, nucleus in an elaboration relation, will most probably contain a term, connected via hypernymy relations with lexical units that belong to the satellite segment. The work on discourse analysis and semantic disambiguation will be partially done within WP4 to support the analysis of portfolios. In fact, in this
Workpackage, we will work mainly on the domain dependent part of this implementation.

The result of the linguistic chains is post-processed in order to have a uniform representation of the annotation in XML. The annotated XML documents will be available in the Documents Repository. The user will have access to the annotated document (the internal elements of the resource) or to the original document, if necessary.

In addition to the linguistic and semantic annotation of a given document, the users will have access to a manual annotation tool in order to enrich documents with user-specific metadata. Users will have access also to social tagging tools via which they can tag documents with tags that are related to a domain ontology or they can create new tags, and relate them to the ontology. We might also think of a way to propose tags to users on the basis of the ontology. The annotation and tagging are vital steps in documents processing, because they will be the metadata that will make the work of a search engine more efficient in finding the most relevant results.

**Image Annotation Service**

The image annotation service can be used to improve the retrievable information of a document. It will be done by annotation of the images in the document with concepts from an ontology. Currently, the annotation of images will be done manually, with the availability of specialized computer-aided support. In future, if a reliable automatic image annotation technology is available, the image annotation will become a part of the annotation pipeline. The process of image annotation is done through the following steps:

- opening an ontology together with a lexicon for ontology translation;
- opening a multimedia document;
- selecting an image;
- annotating the whole image and/or a region representing a particular object in the image with an instance of a concept from the ontology;
- storing the annotated document.

Image selection generally means that the user decides for which document in the document storage and for which image in this document an annotation will be created and stored. The selected image is loaded into the image annotation editor. Then the user can perform a region creation on the image. The region is annotated with one or more appropriate instances. The selection of a concept for the annotation is facilitated by providing shortcuts to the concepts used in the available annotation in the textual part of the document, the annotations of other images in the same document, annotations history, and the domain ontology. Multiple regions in several images can be annotated with the same instance. This means that the same object is depicted in different images. Also, several instances can be used for the annotation of the same region. This means that the region contains a complex object or several objects. For example, if we have ‘a blue iPod’ on one image, we can outline the iPod device on the image in a region and then annotate the region with instances of concepts for 'blue' and 'iPod'. When the annotation of the image is completed, it is incorporated within the document as an addition to the representation of the image within the document. The document is updated in the Document Repository and also in the Index Repository within the XML Search Engine. Thus, the image annotation is made available for user's searching and retrieval.
The image annotation tool is already implemented within the European project AsIsKnown. It has to be integrated within the CSF and, if necessary, the functionality might be extended.

Requirements and tasks

The requirements and the task for the Document Annotation Subsystem are:

- Availability of document converters from PDF, RTF, HTML, MS Word into XML format. If the output is not unified, a post processing conversion step will be needed for unifying the XML formats: XSL transformations or other XML transformation tools can be applied. The CLaRK system is a candidate for this task.

- For the text annotation service, new tools for discourse analysis and semantic disambiguation have to be added to the existing tools from the LT4eL project. Thus, we will reuse the tokenization, lemmatization and concept grammars from LT4eL. Then we will add more elaborating mechanisms for disambiguation of concept information.

- For the image annotation service the main tasks are: the integration within the rest of the CSF, and development of more complex annotations and links to other available annotations for the document.

3.2.2 Semantic Search Subsystem

Description

This subsystem will be used for indexing of and searching for (parts of) documents on the basis of the semantic annotation presented above. After the retrieval, the extracted data can be processed in order to be presented to the users. The basic modules are:

Indexing Management Component

On the basis of the structure of documents the administrator of the CSF specifies the contexts of interest to search in, and determines the dependencies among the separate contexts. Contexts here are defined as elements of the documents. Dependencies are defined via accessibility relations. Both of them are defined via XPath expressions. There is always a root context from which the definitions of the next contexts are defined. When a context is defined on the basis of another context, then it is assumed that the second is directly accessible by the first one. The transitive closure of the direct accessibility relation is the accessibility relation between contexts. Within each context, a set of search items can be defined. Then the search is done on the basis of either the search items for a context, or the contexts that are accessible from it. This way of defining the indexing over documents within the CSF allows us to select contexts with appropriate granularity for a given task.

For example, if we have documents in which the text structure is annotated with structural elements, such as chapters, paragraphs in chapters, sentences in paragraphs, we could use this structure in order to do searches like: give me the paragraphs that contain sentences that are annotated with some specific concepts. In this case, if we have a tutorial, which is about 100 pages long, we would be able to detect the very concrete places which satisfy our query.

After defining the contexts themselves, also their dependencies and the search items are specified for a given type of documents. The actual documents of this kind are analyzed with respect to the defined contexts, dependencies and items. The result of this analysis is then stored in the index repository.
In the case of the semantic annotation, the primary search items are concepts used for the annotation of the documents.

**Query Evaluation Component**

The system supports a query language over search items and contexts. The basic query expressions are single search items (or concepts in this case). Then, on the top of these basic query expressions complex query expressions can be formulated by using boolean operators. In addition to the search items, the queries might contain references to some of the contexts defined in the stored documents. The query is evaluated over the index repository, and the contexts, that satisfy the query, are retrieved.

However, the users of the system will not be able to easily define queries directly in the format, presented above. Thus, we envisage two user friendly ways to define queries. First, the construction of a query expression can be done via navigation over the ontology and the selected concepts to be used as basic search items. The selected concepts from the ontology can be negated or connected with conjunctive or disjunctive operators. With the help of brackets, priorities can be specified. Second, the users specify the query by providing list of keywords. In this case, the keywords are analyzed by the available NLP tools, and then via lexicons are mapped to the concepts from the ontology. The resulting queries in both cases are sent for evaluation.

An additional component here is the query expansion module. Query expansion can be defined as enriching the initial query with information from a given source of information in order to produce a more accurate query for evaluation. Query expansion depends on the type of the search items and their interpretation. In the case of semantic annotation the obvious query expansion is the substitution of each concept with the disjunction of the concept with all of its subconcepts. In our work on the project we will also investigate the appropriateness of other kinds of query expansion.

**Search Result Post Processing Component**

After the retrieval of the appropriate content from the document repository, the user might need to perform some reformatting of the result. This is a task dependent process and it will be defined by the user with the help of predefined XML transformations.

We have already implemented such a semantic search system based on the Lucene search engine. In this project, we envisage to develop the system further with respect to the incorporation of new functionalities for the CSF and better performance at the search phase.

**Requirements and tasks**

- Adjust the Semantic Search Engine to index and search within different predefined types of XML documents;
- Shorten the time for indexing (this is not so critical, thus it would be made as a back process);
- Shorten the time for results retrieval (this is critical. Open questions are how to organize the index repository, so that the search can be made more efficient by prior configuration.)
3.2.3 Ontology Management Subsystem

Description

This subsystem manipulates system resources like lexicons and the ontology. It includes the following modules:

Ontology Reasoning Component:

This component manipulates a registered ontology model and can be extended by importing (registering) other ontology submodels. The component supports the following inference functionalities:

- registration and deregistration of ontology models;
- listing of direct and indirect subconcepts;
- listing of direct and indirect superconcepts;
- listing of individuals for a concept;
- listing of concepts to which an individual belongs;
- listing of properties defined for a concept;
- structural and logical consistency check of the ontology model registered;
- extraction of a registered ontology model;
- generation of ontology fragments (with/out subconcepts; superconcepts; sibling concepts; property relations and range concepts; with property restriction superconcepts for a concept supplied as a parameter);
- generation of a subhierarchy, containing superconcepts and/or subconcepts with a pointed step for a supplied as a parameter concept.

For this module we use the Pellet OWL reasoner.

Ontology Translator Component

This component is used to transform the ontology from its OWL representation into a representation, which is based on a lexicon aligned to the ontology. This task is necessary when the users interact with the ontology. The users could exploit the ontology for some kind of manual annotation, for formulation of a query to the search engine, for navigation over the semantic annotation, etc. Here we rely on the ontology-to-lexicon relation, defined within the LT4eL project. It establishes the connection of concept names in the OWL ontology with the lexical items for the concepts within a language. The component supports the following functionalities:

- transformation of an ontology model into an XML representation with a specified schema. The model is more compact than the original ontology model and contains representation of is-a relations, triples, restrictions on properties (all value, some value, has value, cardinality restrictions).
- translation of the so-transformed ontology model into a specified language.
- translation of an ontology fragment (with/out sub concepts; super concepts; sibling concepts; property relations and range concepts; with property restriction superconcepts for a concept, supplied as a parameter) in a particular language.
- translation of a subhierarchy, containing superconcepts and/or subconcepts with a pointed step for a concept supplied as a parameter in a specified language.

The Ontology Translator Component uses the translation functionalities supplied by the Lexicon Management System.
Lexicon Management System
This system manipulates the lexicons aligned to the ontology in an XML format. It supports functionalities for changing the lexicon content (addition of new lexical items, removal of lexical items) and extraction of data from the lexicon (for a given ontology concept to return the lexicalization for a particular language and vice versa).

Ontology Filtering Component
This component will support the navigation through the ontology in a user-defined manner. It will be based on ontology and lexicon manipulation operations. The operations we envisage for the moment are:

- **Extract minimal ontology**: This operation starts with the selection of several concepts from the ontology. The idea is that a sub-ontology will be extracted. It will contain the concepts and relations which are necessary to support the definitions of the selected concepts. The definition of a concept in these settings is the superconcepts of the selected concepts, the relations defined for these concepts, the concepts necessary for the determination of the range of the relations, superconcepts of all selected in this way concepts. The operation is repeated to reach a fixed point. This operation will be executed on the ontology model.

- **Hide concepts**: If the user is not interested in some concepts of the ontology, he can mark them in the lexicon as not to be displayed. Thus, during the visualization they would not be presented by the UI Ontology Visualization component. Some of the local definitions would be inherited by the subconcepts from their superconcepts and shown, if necessary. This operation will be executed on the lexicon.

- **Hide relations**: The idea here is similar to the ‘Hide concepts’ operation. The user can select some of the relations of a concept and specify in the lexicon that they should not be presented by the UI component. Such relations are not taken into account when the previous two operations are performed. This operation will be executed on the lexicon.

- **Create role context**: This operation will allow users or user groups to create the context of concepts from all concepts that they are interested in. When a particular user from a certain user group is logged in, the ontology view would include just the concepts that are relevant for the particular category and group the user belongs to.

This is an initial list of ontology operations for defining user views over the ontology within the project. However, other possibilities will be also considered.

Ontology Visualization and Navigation Component
Concerning the visualization of the resources that are processed by the CSF, we have the idea to present the material to the user also in a mind map/concept map view. It will support the interaction between the user and ontologies, lexicons. The Ontology Visualization and Navigation Component will use the above services. The visualization will include: tree oriented view; alphabetical view; Mind Map view; Concept Map view; Context view.

Requirements and tasks:
- Shorten time for translation and transformation of an ontology.
- Adapt a Concept/Mind mapping tool for ontology visualization and navigation.
3.2.4 Concept/Mind Map Tool

**Description**

Concept/mind map views are very interactive ways for visualizing trees and graphs. The resources within the CSF are trees and graphs. For example, ontologies are graphs, XML documents are trees. Having in mind these structural peculiarities in the presentation, we propose to illustrate such resources in an easy-to-navigate and locate view with the help of concept/mind maps. We can view the working space for a user or the system repository as an organized set of different kinds of available resources: XML documents, ontologies, lexicons, web repositories. The organization of these resources can be represented as a graph. The basic extensions to the available concept/mind mapping tools would be:

**Ontology Visualization**

The concepts are visualized as nodes in the concept map and the relations between them are named arcs in the concept map (cf. fig. 5). To the nodes in the concept maps other resources can be attached, such as images or internal and external links to data on the local system or the web. In this way, besides the concept name, the node can contain also other types of resource information. The ontology visualization will be always defined via some of the lexicons. The additional resources that can be attached to the nodes can be PPT, MS Word, HTML, PDF documents. They can be directly opened by clicking on the resource. For example, if we have a PDF document that is indexed in our repository and that document has as topic or category some concept from the ontology, the node of the concept can contain a reference to the original document. Thus, we can organize the documents in a document repository using the ontology. The Ontology Filtering functionality will be used to produce an input for the concept map representation.

![Figure 6. An example of an ontology segment visualization.](image)
XML Documents Visualization

The documents that are processed in the CSF can be represented as concept maps themselves. We can define a configuration for different XML schemas. By using XPath pointers, the elements of the document that are required to be visualized as nodes are determined. Also the relations between these nodes are established. In this case, additional resources can be attached, too. For example, some nodes can have a relation to an ontology via which they are conceptually annotated (cf. fig. 6).

Figure 7. An example visualization of related XML documents.

Workspace Visualization

We can assume that the organization of different types of distributed resources and their interrelations can be presented as a concept map. For example, if a user manipulates a set of 3 ontologies, 5 corpora, lexicons in 3 languages, an image repository, he can organize all these resources using a concept map (cf. fig. 7). The main node will be called, for instance, “Workspace”. The child nodes (“Ontologies”, “Documents”, “Lexicons”, and “Images”) will be attached to it. The “Ontologies” node will have one child node for each of the 3 ontologies and each of them will be represented in the way discussed above. The “Documents” node will contain child nodes carrying the name of the 5 separate corpora. These corpora nodes will be root nodes for the documents visualization. The “Lexicons” node will contain child nodes for each lexicon and the lexicons themselves will be presented as discussed before (lexicons are XML documents). Thus we can organize the workspace in a way that is manageable by the learner. The advantage is that we can define additional named arcs between various types of resources. A lexicon can be connected to an ontology node by an arc “translates”; the ontology node can be connected to a corpus node or a document node by an arc "annotates". The Semantic Search Engine, discussed above, would be very useful if integrated, because all the resources in a concept map are XML documents: ontologies, lexicons, learning material and the concept map itself. We can define different indexing schemas and efficient search can be performed within various contexts and with a different degree of granularity. All these steps will
accelerate the navigation process over a big concept map, such as the user's own workspace.

Figure 8. An example workspace visualization.

Requirements:

- Choosing an Open Source concept/mind mapping tool written in Java. We will adapt the VUE tool;
- Formal specification of the functionalities to be developed;
- Development, testing, licensing.

The integration of all these tools as well as the implementation of the missing components will constitute the first version of the CSF which will be further extended with services to support the informal learning. These services are described in section 3.3 - Task 6.2: adding the social component.

3.2.5 Use case 6.1

The design of the functionality for task 6.1 is based on the use case, described in the deliverable D3.1. The use case describes the target group and the context in which the services will be used. It also identifies the stakeholders.

Our main stakeholders are the learner and the tutor. The tutors needs to exchange experience and materials with other tutors and learners. Learners need access to learning materials (with formal or informal origin). Also learners need to share their opinions and to ask to recommendations within their learning community. We aim at developing a framework whose main functionalities are:

1) User-oriented:

A network for interactive and easy to perform user communication should be created so that users can share opinions, comments and feedback with other users. It will include the following facilities:

- support domain specific knowledge models aligned with lexicons (multilingual interactions)
- support filtering of a specific knowledge domain (cultural and social adaptation of the knowledge domain to different users)
D6.1 Social and informal learning support design

- support of annotation, discourse analysis, disambiguation, indexing, metadata creation, LO similarity distance calculation tools (core services that will supply qualified LO creation and manipulation)
- support of unified resource model to which the different resources can be mapped (extensibility of the learning environment with new types and formats of resources)
- support of efficient search engine that can be adapted to different kinds of search within different formats and types of resources
- visualization of resources in an interactive and easy for perception form
- knowledge model for the users network. This model will make possible extensions for universities and their users interaction policies.

2.) Resource-oriented:
The other main components of an eLearning System are the resources. These resources can be of different formats, types and they can cover different knowledge domains. An eLearning system should be integrated with:

- a common manipulation resource model that can describe different kinds of resources in a unified and consistent format
- a number of tools that will process the resources content and make user's access to resources possible
- a number of tools for user's communications and interactions
- conceptual independent models that will define a common conceptualization for the different scientific domains (domain ontologies)
- knowledge model for the resources network. This model will facilitate the adapting of new resources formats and types as well as their manipulation by the available tools.

3.2.6 Showcase 6.1
The integration of the various tools and resources developed within the LT4eL project into the Common Semantic Framework is a part of the showcase (see D3.1) for this task. The extension of the functionalities with a more precise annotation of the learning objects and more accurate navigation over the various resources will provide a better environment for the formal learning related task. In the initial phase, we test the CSF at the technological level. It is performed by a comparison of the results from some predefined search queries. We will also perform testing with tutors who would use the CSF for the creation of a new course. This course would include the list of the appropriate learning objects as well as some educational recommendations to the learners.

3.3 Task 6.2: adding the social component
Learners need to be guided in their learning process and this can be achieved by means of ontologies which provide a formalization of the domain knowledge provided by experts. On the other hand, social sites such as the ones mentioned in section 2 often include tags that are assigned to the various resources by users on the basis of their understanding of the resources and might even include their appreciation for the material proposed.
The Common Semantic Framework, after inclusion of the services by task 6.2, will be a knowledge sharing system that connects learners to resources and learners to other learners by means of user profiles, ontologies and social tagging. In order to connect
learners to resources, we relate the tags that emerge from the social media applications to the (lexicalization of the) concepts of an existing domain ontology. To this end, we extract the domain knowledge (Task 6.2.1) based on social tagging (Task 6.2.3). This domain knowledge, covering relations between topics is compared with that attested in ontologies and is eventually used to populate them (Task 6.2.2). In this way, we aim to complement the formal knowledge represented by ontologies with the informal knowledge emerging from social tagging improving thus the possibility of retrieving appropriate material for a more personalized learning experience.

We also aim at connecting learners to other learners. To this end, the content the learner is searching and selecting can be used as a trigger to get him in touch with other users who have tagged this content or used this content before him. This will be the case if the learner is a novice and needs to create his own community with people with similar interests. Alternatively, if the learner is part of an already established community based on common interests he will need to be updated with the changes in his domain(s) of interest. To this end, the learner will focus on the learning objects that are produced by people who are relevant for the domain he studies and/or people the learner trusts. The system will monitor the changes that appear in his network with respect to content and to users and will recommend how the learner should update the network (by adding new peers or removing old ones) and answering the learner’s questions about relevant materials and peers on a given topic. In this way, we add a trust dimension to the search since a learner will trust the objects produced, tagged or recommended by his own network.

In this system, the figure of the teacher/tutor might change dynamically and might not always be a unique figure but a series of users that can satisfy his different learning needs. From this perspective, a tutor can be any content generator in the network and the learner rates the tutors on the basis of his appreciation for what the tutor writes or recommends.

The following figure visualizes the most important flows of data related to task 6.2.
The following services will be realized within the context of task 6.2 and will be integrated into the CSF:

**Tagging Service and Tag Analyzer**

The system will ask the user to tag:
- learning material and other resources
- their own contributions
- other users’ contributions

In addition, the system will import tags from existing social sites for similar or related resources which are identified by a unique resource identifier.

**Ontology enrichment feedback**

After a user provides tags, those tags must be stored, and matched to concepts, users and resources. There will be a service that does this with the help of the user. For each tag the user has provided, the system selects related tags and/or concepts and asks him whether this is what he meant by this tag. The user can choose between “yes”, “similar/related” and “no”. If the user selects “similar/related”, the system gives some options to indicate in what way the tag is similar/related to the proposed one. We are
also investigating ways to make this process automatic by relying on the existing knowledge already attested in social sites and ontologies. Furthermore, when a relation between two tags is derived from the tagged data, this relation can be transferred to the ontology. However, it is unknown what is the type of relation (e.g. PartOf, UsedBy). This info can be acquired by asking the learners.

Social Network Assistant
This service will discover the learner’s external online social network, and use information about topics, potential learning materials and users of other social media for the recommendation service. More details are described in section 3.5.

Material and person recommendation
The learner will be pointed to resources that are probably relevant, and to other learners that might be able to help and advise. The system will make use of knowledge external to the system, and user information from other social networking communities.

Rating service
In order to collect opinions from learners, the system will have a rating service for resources and feedback he has received assessing the general quality of the resource/feedback for the task he is carrying out. Since we will require that each learner assigns tags to the resources he employs, the validity of a given tag for a given resource will be assessed on the basis of how many people have tagged the same resource with the same tag.

User history service (not in picture)
There will be a user history service that keeps track of users’ interactions with the system, and uses this information to suggest resources that are relevant for the user. It can either be used to filter out irrelevant information, or to higher rank the relevant information. Two kinds of information can be used:

1. Interaction history of the user in question, to determine his interests, competence, and similarity between user’s profile and a given resource.
2. Interaction history of all users, to see which information they used for a topic or task similar to the current situation for the user in question.

Points Service (not in picture)
In order to support shared knowledge, learners should also contribute to the knowledge in the system. An extrinsic motivation to do this will be created by rewarding people who contribute with points, where a high number of points gives them more status. An even more restrictive system would allow users only to do certain actions (e.g. post a question to a forum, view learning materials) if they have a positive number of credit points, which can be earned by contributing to public knowledge. The service calculates how many system credits are earned or deducted for certain contributions and actions (such as providing tags). A similar system is described by Artail et al. (2008).
3.3.1 Use case 6.2

The design of the functionality for task 6.2 is based on a use case, i.e. Use Case 6.2, described in the deliverable D3.1. The use case describes the target group and the context in which the services will be used and identifies the stakeholders.

The target group, lifelong learners, is a heterogeneous group: they have different backgrounds, different ages etc. These learners are both specialists and non-specialists, who have to acquire certain skills and knowledge or to keep up their skills and knowledge with the new developments that appear in their domain of expertise. Their learning topic is on a domain where there is a lot of information on the web, e.g. computing. The learners do not have a face-to-face course at their disposal. However, they want to or have to receive a grade or some kind of proof that they acquired the knowledge/skills. This confirmation is provided by the organization that has an interest in the learning process or by the community in which the learner interacts.

The organization can be an educational institute, but also a company that wants their employees to acquire certain skills and knowledge. The learners and teachers/tutors are also part of the knowledge sharing community described in UC6.1.

The main stakeholders in our task are the learner and the tutor. The following problems have been identified for those stakeholders:

**Learner:**
1. Is looking for a detailed answer to a specific question and cannot find it in the material that is currently available to him.
2. Wants to find out which topics are related and which of them might be relevant to carry out his learning task.
3. Wants to know which resources to access in order to be well-informed about a certain topic.
4. Wants to find people who can help him (e.g. to give feedback on produced work or to recommend useful materials). Those people can be:
   - Persons in his current networks on social networking websites.
   - Unacquainted persons that take the same course or have the same learning objectives.
5. Wants to monitor the activity in a domain in order to be always aware of the state of the art.

**Teacher/tutor:**
1. Needs to monitor which knowledge and skills a learner has acquired (progress monitoring).
2. Wants to know which resources learners are using.
3. Wants to reuse information which emerged from the learning process, such as questions from earlier students, feedback given to them and information about which materials helped them to accomplish certain tasks.

The services developed in the workpackage will provide support to solve the problems listed in the problem description.

**The learner will be able to:**
1. Search for materials in several ways, as described in UC6.1.
2. Get recommendations for learning material on the basis of user profile, social network and history.
3. Access meta information about resources, given by other learners.
4. Get recommendations about users with knowledge/skills relevant for the current task.

On the other hand, tutors will be able to:

1. To monitor to some extent what the learners are doing and have acquired.
2. To see which resources are accessed through the system. However, it is not realistic to monitor whether and to what extent they are really used.
3. To inspect accumulated knowledge available in the system, such as ratings and tags. This information can be used to help current learners.

3.4 Task 6.2a: domain knowledge & social tagging

This section deals with one aspect of "adding the social component", namely the content side, with respect to knowledge in the form of social tagging data and ontologies. We call this task 6.2a, while in task 6.2b the focus is on individual users and social networks.

We will first describe in which way social tagging will help us and we describe a way to extract relations from various media. In particular, we conducted an experiment on data from Delicious. Finally, we discuss the interaction between ontologies, folksonomies and learning material.

The following figure gives a graphical representation of the sources and processing of data for task 6.2a.

Figure 10. The sources and processing of data for task 6.2a. The left side of the picture displays the approach for finding related tags, based on data retrieved from social tagging systems. This is
an input for ontology enrichment, as well as relations from Wikipedia and keywords from other sources that are not tagged.

### 3.4.1 Social Tagging

As part of the services to be developed, we will provide a social tagging feature for users. This is useful in a direct and an indirect way. A user can directly benefit from providing tags, since it helps him find back his resources and remember what they are about: the tags function as labels, categories and summaries. But at the same time, from social tagging data, certain information emerges, such as:

- Which resources are popular? (They have tags from more users than other resources do)
- What is a good tag for a certain resource? (Tags that are assigned to this resource by many users)
- Which resources are about the same topics? (They have the same tags)
- Which tags are related? (They occur together at several resources)

Those are thus the indirect benefits. The service that recommends materials to users will make use of this information.

In the following sections, we focus on the latter question (which tags are related), in order to relate the underlying tagging knowledge to formal knowledge.

### 3.4.2 Relation extraction methodology

In section 2, we mentioned some approaches that try different algorithms, such as cosine similarity, for clustering of tags and finding relations. Since in this project, we deal with many practical aspects of the social dimension in learning, we start with just a simple approach for finding relations between tags and focus on combining it with other kinds of knowledge rather than on implementing complex algorithms. Once everything is in place, it can be judged which aspects or components are most worth improving.

We therefore investigated an easy way of extracting related tags, based on co-occurrence on a set of bookmarks. We look at the knowledge from a perspective where the user’s characteristics are hidden: it is not relevant which other resources a user has tagged or what his user profile looks like, but only the number of different users who tagged a resource with a certain tag is taken into account. From this perspective, social bookmarking is seen as a way to discover which words (tags) are good descriptions for a certain resource, rather than a way of relating users to materials. The knowledge is still valid if the user accounts have ceased to exist, assuming that the resources have not changed that much.

The retrieval of related tags is based on the following assumptions:

- When topics are closely related, they will often occur together in the same resource.
- The most important topics of a bookmarked resource are represented by a tag for the bookmark, given by several users.

The following main methodology was deduced from the experiment with data from Delicious as describer further below:

- For a seed tag, retrieve bookmarks that have this tag in their top-5 of most-used tags (exclude bookmarks that are mainly about different topics than the seed tag).
• Per class of bookmarks, a set of candidates for related tags is created by adding from its bookmark its 10 most applied tags, but only those tags that are given by 2 or more users, or 5% of the number of savers of this bookmark if this is more than 2.
• Related tags are found by selecting tags that are assigned to at least a certain percentage of bookmarks in a class.

The following parameters influence the quantity and quality of the found relations:
• number users that tagged a certain bookmark
• size of the set of bookmarks
• number of selected most-used tags
• minimum number/percentage of users for a tag-bookmark combination (we found that 2 users or 5% works)
• minimum number/percentage of different bookmarks for which a tag must co-occur (also dependent on size of the set)

3.4.3 Experiment with sparse, anonymous data from Delicious
In an experiment, we made use of data from the social bookmarking website delicious.com, where users store their favorite URLs with tags. With around five million users, Delicious is one of the biggest social bookmarking / social tagging websites on the internet.

In particular, we address the following:
• whether it is possible to find tags that are related by co-occurrence in case only few users have tagged the relevant resources, and in case only few tagged resources are available. This is important to assess, because if we apply social tagging to learning material within an eLearning system, we anticipate situations with few users and few resources.
• how the found related tags correspond with concepts and lexicalizations that are already present in the LT4eL ontology.

The methodology is the following (part of this is also chosen as the general methodology described above):
• Select start/seed tags that serve as the basis to retrieve tags related to them. Choose tags that are in the LT4eL ontology but that are also popular as tags on the social tagging system.
• For a seed tag, retrieve bookmarks that have this tag in their top-5 of most-used tags (exclude bookmarks that are mainly about different topics than the seed tag).
• Categorize the bookmarks by classes based on how many users have bookmarked it, e.g. 14-25 users, 26-50 users.
• Per class of bookmarks, a set of candidates for related tags is created by adding from its bookmark its 10 most applied tags, but only those tags that are given by 2 or more users, or 5% of the number of savers of this bookmark if this is more than 2.
• Related tags are found by selecting tags that are assigned to at least a certain percentage of bookmarks in a class. Also, we experimented with the size of the sample of bookmarks.
Gold standard
Delicious offers a list of the top-11 related tags for a given tag. This shows which tags generally occur together on Delicious. We consider it as a kind of gold standard, and compare the related tags found by the small sets of bookmarks with this list.

Findings
We experimented with sets of 10, 15 and 75 bookmarks, and several classes of numbers of savers of a bookmark. The different sets each yielded 3 to 6 of the 11 related tags from the gold standard. For the sets of size 15 or more, no false positives were found. Thus it appears that if topics are related, some of them also do occur together in a relatively small sample of bookmarks. A bigger sample can help to reduce the “randomness effect” of the selection and reduce the number of false positives: the correct relations can still be found while increasing the absolute number of bookmarks required for co-occurrence, and the tags that happen to occur on only 3 or 4 bookmarks are ruled out.

The number of people who tagged a bookmark has an effect on recall: the classes with more people yielded more related tags. However, we can conclude that bookmarks tagged by around 10 people are a valid resource for deriving related tags. With higher numbers of people, a few more relations can be found, but the results are similar.

This means that we can start analyzing and using accumulated knowledge in the system we are going to develop once the first users start adding and tagging resources. Also, the methodology can be applied to tagging data from other systems, if the tags for a resource are sorted by frequency, so that the top-5 can be determined.

3.4.4 Applying the approach to other knowledge sharing platforms
From the inventory of knowledge sharing communities in section 2, we see that different kinds of resources are shared, and not all sites use tagging. Some sites show related tags, not always specifying how they calculate relatedness. The relations can be based on the combination of interests among multiple users (several users use both tag A and tag B), or on the tagged material (several bookmarks are tagged with both A and B). Both are described by Mika (2005) on the basis of data from del.icio.us: the graph of all the relations is called a light-weight ontology.

Our approach described above takes into account the relations based on the tagged materials. In principle, it can be applied to extract relations if there is a collection of items that are tagged with multiple tags.

However, retrieving a relevant collection is not always as easy as with Delicious, which is an excellent resource for this purpose. Some social media applications that do offer the possibility to tag do not allow to search using tags, and all social media applications offer the possibility to store resources without attaching a tag. This restricts the search scope. In Shelfari for instance, if one looks for "javascript", one retrieves a list of books with "javascript" in the title. Subsequently, one can view the info for each of those books to see whether there are tags that can provide information on what else the book is about, e.g. "webdesign". However, one would not find a book called "Webdesign" even if it has the tag "javascript" attached. Furthermore, a search can result in many items that match the search word in a title but do not have any tags. In YouTube, it is the other way around: the owner of a video can enter tags so that the video will appear in search results if other users search for those words, but the list of tags for a certain video is not visible.
D6.1 Social and informal learning support design

Obviously, not all kinds of resources are equally suitable for learning or for the domains of computer science, e.g. music ratings from last.fm do not match our project.

Retrieving related topics from Wikipedia

Similar to the procedure described above for finding related tags through co-occurrence on bookmarks, related topics can be found in pages on Wikipedia, even though no social tagging is involved. We can make use of the fact that every Wikipedia article is dedicated to one topic, and the word or phrase best describing the topic is the title as well as the link to the article. Thus, the relations that can be retrieved are those between the title (current topic) and each of the links to other Wikipedia pages (related topic). This idea is also exploited by www.wikimindmap.org, where a mindmap view of a Wikipedia page is given. The relations reflect collaborative knowledge in the sense that a Wikipedia article is usually the result of a collaborative effort: if an author forgets information or includes wrong information, this is corrected by other authors.

The relevance of the relations between topics found in this way can be justified because two criteria hold: a) If the related topic would have nothing to do with the current topic, it would not be mentioned in the article. b) If the related topic would not be relevant as such or in the context, there would be no internal link to the corresponding Wikipedia article.

A danger of this method is that a huge amount of related topics can be found. E.g. the Wikipedia entry for "Linux" contains around 240 internal links. So an additional way of filtering is needed. This can be done by comparing all those topics with another information source, such as a user profile or ontology, or by making a selection based on the Wikipedia article itself. For example, a possibility is to consider only the links in the introduction of the article.

If we do this for the topic "linux", we get 16 topics, 3 of which are equivalent with a term from the top-11 related tags given by Delicious. So the information from Delicious can be a useful addition to related topics retrieved from social tagging data.

An additional way of exploiting the relations can be obtained if relations are extracted and subsequently used in two directions: besides the way described above, also topics can be found whose Wikipedia article contains a link to the current topic.

3.4.5 Interaction between ontologies, folksonomies and learning material

By employing the principle of social tagging in an eLearning community, the knowledge represented by the tags can be used for several purposes, among which:

- Retrieving learning materials directly or through the emerging folksonomy
- The ontology that is used for searching learning material can be enriched with the found terms/concepts and relations

In the next phase of the project, we will investigate in more detail to which extent the combination of folksonomies and a domain ontology is possible and useful.

In our experiment, most found related tags are in the domain of computer science. From those tags, around 50% are in the ontology. Both those observations are determined by human assessment; for the latter, a mapping of tags to concepts is performed manually. This means there are at least two challenges:
Automatic mapping of related tags to existing concepts if applicable (around 50% of found tags)

Insertion of new tags in ontology (the remaining 50%)

We can add the found relations between two tags if both are mapped to existing concepts. The figure below is an example of a part of the ontology with the relations from "tools" to related tags indicated by arrows. The relations are in one direction. E.g. there is a relation from "tools" to "software", indicating that together with "tools", one of the most frequently co-occurring tags is "software". The other way around, however, starting from "software", "tools" does not necessarily co-occur as often with "software". This difference is caused by the fact that for each seed tag, a different collection of bookmarks is used, that represents the seed tag well, so each relation is derived from a different collection. The relations would obviously be symmetrical if co-occurrences are counted for one fixed set of bookmarks.

![Figure 11. An extract of the computer science ontology of LT4eL, supplemented with relations derived from Delicious data: the arrows mean that together with the tag "tools", the following tags frequently occur as well: software, windows, utility, web, tool. The figure shows, where in the hierarchical structure of the ontology they can be found.](image)

We see many relations between concepts that are far apart in the taxonomy. This means that in the non-enriched ontology, when starting with one of those concepts, the other one would neither easily be found by manually navigating through the ontology nor by simple expansion with parent or child nodes. Relations derived from social tagging data can thus be a useful addition, in order to bring concepts closer together than the hierarchical structure offers. It is not clear, what is the type of relation, other than "co-occurs with". However, probably it suffices to give a user the option to select a related concept, or to expand a query with a related concept, even if the type of relation is not specified. Most users will make the link between two concepts in their minds immediately, without the need to read what is the relation. On the other hand, relations can play a role in automatic reasoning.
Some approaches describe how relations between concepts can be found using online resources. Specia and Motta (2007), where tags are clustered to terms describing concepts, and relations between concepts are looked up in existing ontologies, while Van Damme et al. (2007) describe steps to turn a folksonomy into an ontology.

We envisage additional ways of determining the kind of relation between two concepts. A relatively simple one is to use ConceptNet (Havasi et al. 2007): a database with commonsense knowledge, stored as concept-relation-concept pairs. Unfortunately, this approach seems less promising for the domain of computer science. A first inventory shows, that most of the popular tags about computer science are not represented in ConceptNet.

A high-precision solution is to ask the user to choose from a predefined set of relations, as part of the knowledge collection process where users earn credit points. If multiple users are asked, this is a way of collecting collaborative knowledge as well.

Furthermore, we can look whether the relations can be extracted from textual learning materials that have been tagged with the regarding terms. This idea can be extended by concordance search in domain corpora.

### 3.4.6 Showcase 6.2a

The experiment with tagging data from Delicious is part of the showcase (see D3.1) for this task, in which we will furthermore assess to which extent it is possible to enhance ontologies semi-automatically on the basis of social tagging. We will act as experts to evaluate the found related tags and their appropriateness to enrich an ontology - in particular, the ontology on the computer science domain from LT4eL. Furthermore, we will assess what are the advantages and disadvantages of enhancing ontologies with social tagging for searching and suggesting relevant material to both beginners and advanced learners.

### 3.5 Task 6.2b: social networks

In section 2, we have established that the individual learns by being in a community of practice and evolving inside that community. Also the individual learns influenced by the social environment which he is part of. Our task here is to identify the communities that the user is part of, to help him manage them, to help him evolve inside the community, by creating and improving valuable network connections.

In order to do that we have analyzed the behavior of the user inside some social media communities (YouTube, Delicious, blogger, etc.) and we have discovered that there are 2 main types of cases that we should consider:

- The community is strictly contained inside the social networking application. The relations between users are only the one provided by the network. These relations are easier to extract and analyze as they are explicitly provided.
- An individual uses multiple social media applications or one community uses more social networking applications to better organize their activity. In this case the relations between users are spread across multiple social network sites and the key to extract and analyze these relations. In order to do that an important task is to identify the users who are part of multiple social network applications because they represent the bridge between those specific communities.

We also have to define more specifically what we understand by the term community.
The users that are using a social network application (YouTube, Delicious, etc.) form a community. This community is highly heterogeneous because it is formed by people with many interests. The centrality aspects of the individuals in one of these communities is not very relevant from the learning point of view because the centrality isn't linked in most of the cases with the type of content a specific user is interested in. A user being central in such a community shows dedication to it and might indicate good communication and networking skills. In the rest of the document we will refer to this type of communities by social communities.

A community is also formed by the users that share some specific interest. The more specific the domain, the smaller and more specialized is the community. For example there is a huge medical community and there is a smaller community interested in cancer and there is a smaller and more specialized community interested in lung cancer. The community focused around a more specific topic is included mostly in the community formed around a more general topic. A person that is central and therefore considered an expert in the larger community might be a novice in some of the more specialized ones. Also these communities might be using different social network applications for knowledge sharing. These communities will be referred to in the rest of the document as specialized communities.

As we have said in the beginning, a user can be part of one or more social communities and can also be part of one or more specialized communities. In case he is part of more communities the individual represents a bridge that can help people with different interests to be linked together based on a possible common interest. Our research will be focused on discovering these persons.

The main issues are

- discovery of the user's social network
- management of the user's social network
- analysis of the user's social network
- search and retrieval of content from the network

We will analyze how we can perform each of these tasks and then we will present the design of the framework that will perform these tasks.

### 3.5.1 Discovery of the user's social network

The user provides a list with his usernames on the social media applications and the communities he is part of.

The most relevant for us will be the blogging platforms, social bookmarking, content sharing (YouTube, Flickr). A specialized crawler will extract the relations between the usernames given and their peers, as well as the content produced and recommended by the peers. The data extracted will be actually metadata describing the user's network in terms of relations and in terms of content.

The discovery will be done considering the following ideas: On each of the social communities, the list of peers of the user is extracted until a given degree. The degrees of separation between two individuals is represented by the minimum number of connections that separate them. For example if X and Y have a mutual friend Z but they don't know each other then X and Y are separated by 2 degrees (the connection between X and Z and the connection between Z and Y). We are not interested in monitoring the users that are too far, as it has been established that even if we can be linked with everyone on the planet through 6 or 7 degrees of separation, most of the
second degree links and almost all of the third degree links are very difficult to use. Also the type of relation between peers should be identified at this step.

- **Main goals**
  - Identify social connections inside social communities
  - Identify possible peers across social communities
  - Identify and store metadata referring to the data produced inside the user's social network

- **Inputs**
  - The list of social network communities used by the user and the usernames to access them

- **Outputs**
  - Tree of relations for this user
  - List of content generated inside the social network
  - Triples of type (user, content, tag) that will be used for analysis, search and recommendation

We have identified 2 techniques to be used:

- **Feed analysis** - most of the social network applications provide web services regarding the content and some of the relations between users. These services are mostly provided through feeds. The problem here is that generally the services do not provide the relations between users, the comments and the full data about the tag usage.

- **Crawling** - the rest of the necessary data can be extracted through crawling when available. This task has also to consider aspects like not stressing the social networking applications. That is why a caching mechanism needs to be implemented and also the data needs to be fetched from the user's computer. Considering these aspects we need to develop a distributed architecture for data extraction.

### 3.5.2 Management of the user's social network

Research shows that we generally keep connections with about 150 people. The people we know through social communities can represent a much bigger number. The challenge here is to present the information to the user, offer visualization tools and offer recommendations on how to extend the network and on how to keep some important contacts.

The main goals here will be:

- **Offer advanced visualization tools that will allow to**
  - Identify which social communities a user is part of
  - Identify which specialized communities a user is part of
  - Identify how important a user is for a given community
  - Check if the relation is properly maintained (the level of communication between users) - this is very difficult to monitor as communication can be performed on channels that cannot be examined by our system

- **Offer recommendations**
  - Recommend adding users that are important for some specialized community and should be monitored
  - Recommend eliminating users that do not provide contributions or that are spamming
3.5.3 Analysis of the social network

The user will be presented with the most important indicators for the social network and also with explanations for them. These indicators are:

- **Centrality** - the total number of ties to other actors of the network. The centrality measures the contribution of the experts inside the community to the learning process.
- **Centrality eigenvector** - this measure also considers the importance of the peers inside the social network. It computes the eigenvector for all the nodes in the network so that like the pagerank algorithm the importance of a node is considered when linking to another, thus meaning that a relation between two experts is more relevant than one between two novices.
- **Cohesion** - this will basically measure the number of cliques (complete subgraphs) inside the network. It is likely that the cohesion is bigger for the more specialized networks. This aspect remains to be studied.
- **Radiality** - “the degree an individual’s network (formed by individuals he knows) reaches out into the network and provides novel information and influence” (Valente 98). This will be probably bigger for the social communities or for the very general social networks.
- **Network reach** - is basically the set of peers that are connected to a user until the 2nd or 3rd degree. This is what we extract in the discovery stage. It can show if the individual is well connected to all the important persons in the network.
- **Density** - represents the cohesion inside a network. It can be used to differentiate the more specialized networks from the others if we determine that the cohesion if bigger in the more specialized networks.

These indicators together with an interpretation can show the user relevant information about the networks he belongs to.

3.5.4 Search and retrieval of content from the network

This part will use the data collected by the network discovery module in order to obtain certain rankings of content based on user preferences. In other words, this is a social network-based search engine. The advantage of such an engine over a classical one would be the level of trust of the results. These results are returned from what people in one’s network produced or recommended through tags, therefore are trusted by the user.

The algorithm that we are going to use will rank the documents or users according to some keywords entered. The ranking can be done based on a group of words provided (representing possible tags), a name (representing possible other user) or a document. The algorithm used is a modified version of FolkRank algorithm. The original one is described in (Hotho et al., 2006).

The algorithm works in the following way. First of all a matrix of the social network graph has to be created. Each user (name), document (resource) and tag (word) Each tuple of type user-document-tag increases the weight of the corresponding user-document, document-tag and user-tag edges. So, in the end, the weight of one edge represents the number of tuples in which that pair appears. This is, a relation between a user and a tag for example is stronger if there are more resources belonging to that user tagged with that tag.
Afterward, each column of the matrix is normalized. Following, this matrix will be named A. A is a n×n matrix, where n is the total number of users, resources and tags in the network of the user of this application. Denoting by w the (final) weight vector (w[i] – the score of the user/resource/tag i), by p a (normalized) preference vector (depending on the type of the search and the words provided) and by d a coefficient from [0;1], through multiple iterations of type \( w \leftarrow dA w + (1 - d)p \) the values in w will get closer to a fix point value.

Using two different values for the coefficient d, we obtain w0 (for d=1) and w1. Although in the original algorithm the final score is the difference between the two values we proposed to use a pondered difference, in thought that we simply must eliminate from \( w1 = dA w + (1-d)p \) the influence of the red term witch can be done in some way by subtracting \( d\times w0 \) in order to obtain a better ranking based on the network (the matrix A is in the formula) but reflecting the preference of the user (otherwise it would have had the same result each time).

A very big importance has the preference vector p. Various scores have been given to the terms on different criteria, including synonymy deduced from FreeBase. We used Freebase for measuring similarity between terms instead of Wordnet as there are a lot of concepts in our domains - like names of technologies for example that cannot be found in Wordnet. We also propose to modify the d coefficient in order to emulate the real particularities of the three types of searches.

What we are going to study and try to improve are the following issues:

- Identify persons who tag without meaning (spammers) and isolate them. We suggest that a way to do this would be to reduce the importance of objects that have too many tags, considering the users didn’t know to use tagging. Also a resource with many tags is of great generality, and perhaps it isn’t interesting for a user searching a particular thing. For this to be done modifying the construction of matrix A - inverting certain element \( x \rightarrow 1/x \) or transforming d in a n-dimension scalar coefficient must be considered;
- Improve the computational performances of the algorithm (as in time and space used, not results); the number of steps to iterate in computing w can be a point of interest;
- Finding relations between search words provided and existing tags I database (not finished);
- Calibrate the coefficient d, by multiple tests using real networks.

Most importantly this search could be improved if the similarity measures that we use were actually based on an ontology. We use for the moment Freebase as a substitute for an ontology as it provides (limited) formal relations between concepts. An ontology of the domain should overcome the limitations of Freebase and we expect the results of the search to be improved by a significant margin. This should be done by linking the tags to the concepts in an ontology and use the ontology to compute the similarity between tags instead of using an external knowledge base that it is not specialized on our domain of work. As the ontologies of the domains will be enhanced with tags in the way previously described, they will be integrated with the search engine.

3.5.5 Design of the framework and services

The current design consists of three distinct modules:

- The data extraction module
The search module
The user interface module

The design of the application is described in the following picture. The modules are distinct and can be implemented in a language of choice. They will provide results to the interface level through web services.

Figure 11. The architecture of an application that can support a user's social learning activities. The image presents the most important modules of the application.

The crawled data consists in a repository stored on the user's computer. This data will be used for assisting the user in search and with recommendations and it will be linked with the semantic framework and with the front end through web services.

This repository is a database that contains the knowledge referring to the user's social network. The data could be serialized in a way that allows interoperability like FOAF or SIOC or it can be used for personal purposes like search, recommendations and management. The preliminary structure of the database model is presented in the
The model focuses on the triple (User, Resource, Tag) which is the way to represent this type of knowledge according to Mika (2005). The rest of the concepts stored in the database are about the users' communities, a queue that is used by the discovery tool and a cache containing information about the senses of the words analyzed.

The main idea is to gather data about the social network of a user across the different communities that the user is using. On this communities the user can create, tag, read, bookmark resources. The user has relations inside these communities with various other users through different types of relations: fan, friendship, subscriber, blog reader, friend as described for example by the XFN standard.

In order to keep track of these activities the above Entity-Relation model is used. The main entities are:

- User - which represents the user of the application and his friends
- Resource - which represents any type of content that a User can create or use (videos, bookmarks, blog entries, etc.)
- Tag - a word that is used to annotate a resource
- Community - a virtual group of users where a number of tags are used

Auxiliary entities are:

- Queue - the list of urls to be visited and the next time of the visit - this is done considering the fact that the content relevant for a user is not very often updated and we don't want to pressure to much the content providers.
- Other entities might be added in time

Between the main entities there are a number of important relations:

- User-user : many to many relation that represents the type of relation existent between 2 different users. The type of relation is an attribute of the database relation. The user-user relation is discovered through the discovery module.
- User belongs to community - the relations is built when a user is discovered on a specific community (YouTube, Delicious, etc)
- User to resource - many to many relation that represents the type of relation between a user and a resource
  - User can be author of the resource - in case of blog entries
  - User can tag or bookmark a specific resource
  - User can comment upon the resource (video on youtube, blog entry, etc.)
- User-resource-tag - this relation identifies the triple identified in Mika (2005);

This model as well as the architecture of the application can be modified during the project considering the further integration requirements.

3.5.6 Preliminary results

Considering the architecture and the database model designed we have performed a series of experiments using the ideas presented in this document.

We have developed a crawler for youtube.com and we have extracted the data started from a popular user and going 2 levels deep. We have discovered 826 users, 9684 resources (films in this case), 17253 tags and 98900 tag associations.

We have implemented our version of the FolkRank algorithm presented above and we have tested it for the data we have already gathered. Also a preliminary interface for the user has been developed.

The experiments showed the following:

- The graph containing the users, resources and tags is mostly connected. It has
  - 54 components with 1 node
  - 1 component with 27706 nodes
  - 1 component with 3 nodes

This means that we can determine the w1 array that we use in FolkRank in an unique way (because we have a connected graph)

- The fact that each community has its own language is very easy to observe by analyzing the clusters of tags. This clusters will contain words that usually are not semantically related but can be related in the context of a community. An example of different semantics according to the community is:
  - user1: [eclipse] [Robert] [kristen] [pattinson] [Stewart] [twilight] [trailer] [moon] [cast] [meyer] [preview] [stephenie] [vampires] [official] [movie] [new] - eclipse as an astronomic phenomenon
  - user2: [eclipse] [Jopen] [1:32] [clubjapo] [proyecto] - eclipse as a music related item
We have applied the Markov Clustering Algorithms (van Dongen, 2000) to identify groups of related tags. We have considered that 2 tags are related if they used simultaneously by two or more users. This will also solve another problem we have detected - the one of the tag spammers. The tag spammers are users that tag a resource with many concepts some of them being irrelevant. One of the most common spamming is tagging a resource with the user's username. This is unnecessary for a search engine and can induce erroneous relations. Tag clustering can help improve the search by identifying at first groups of words that are related.

On this experiment we have identified 509 clusters containing 1475 words. The first 5 of them are the following:

1. obama 2008 mccain Election president barack Hillary Clinton convention DNC democrat campaign primary presidential elections voting vote Republican Democratic
2. metallica hetfield kirk lars ulrich else matters
3. vista windows apple PC macintosh jobs
4. soldier war gears vietnam troops israel
5. iraq oil Iran Soldiers torture policy

This experiment didn't focus on the computing domain but rather on random users. This was due mainly because we wanted to test the discovery module on a large network and this one was one of the largest and because we lacked for the moment the integration with the formal ontologies on the computer domain.

The preliminary conclusions of the experiments are:

- we have designed a knowledge discovery and acquisition tool that can bring valuable information to the users
- we have identified, proposed improvements and implemented search algorithms in order to facilitate knowledge retrieval from the social network.
- we have to propose algorithms for recommending peers both inside and outside the network. These algorithms can be based on the existing work

The most important aspect of this architecture from our point of view is that the user does not need to enroll in a specific social community to get feedback from his/hers social relations. This might succeed in getting the user to actually use the software - one of the biggest problems in the domain of computer supported collaborative learning.

### 3.5.7 Showcase 6.2b

The showcase 6.2b is based on the social network management application that is described in this section. The application will be developed and tested in the Politehnica University of Bucharest using the Computer Science domain. The showcase will aim to show the way the users are using their social learning network and will show them the way of improving this usage. Some indicators of the actual usage of the social network will be computed and analyzed.
4 Conclusion

In this first phase of the project, we have established the basic requirements for the Common Semantic Framework which should support formal and informal learning.

To this end, a review of the state of the art has been carried out which includes an inventory of possible tools and methodologies that can be employed for its implementation. In addition, the most important aspects of the informal learning component, that is social media applications, tagging and social networks have been assessed. The goal of this study was to identify the appropriate methodology and technology for the integration of the formal and informal learning components within the Common Semantic Framework.

A preliminary architecture of the Common Semantic Framework has been proposed and the various services that will be developed have been specified. The research has also focused on how to extend the functionalities developed in the LT4eL project to deal with formal learning and to assess which revision and adaptation are necessary for our purposes. In addition, we have investigated the way we are going to relate informal material, tagging and communities of learners to formal learning material which has been structured by means of an ontology. Experiments have been carried out to this end and have been summarized in this report.

In this first phase of the project, we have worked at the formal and the informal learning components separately and we have investigated ways to connect the two. In the second phase, we will start with the implementation of these various components and will focus more specifically on the integration design.
5 References


Appendix A

This Appendix contains more detailed information about the operational modules that are overviewed in section 2 “State of the art”.

<table>
<thead>
<tr>
<th>Tool Name</th>
<th>Description</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>wikimindmap</td>
<td>wikimindmap is a tool to browse easily and efficiently in Wiki content, inspired by the mindmap technique. Wiki pages in large public wiki’s, such as wikipedia, have become rich and complex documents. Thus, it is not always straightforward to find the information that one is looking for. This tool aims to support users to get a good structured and easy understandable overview of the topic one is looking for.</td>
<td>The tool is developed for wiki content browsing and structuring. Other resources cannot be organized for browsing and organizing.</td>
</tr>
<tr>
<td>freemind</td>
<td>Freemind is a premier free mind-mapping software written in Java. The tool supports: browsing the local file system as a map, creation of a mind map from scratch, addition of local and remote hyperlinks to a node, merging of maps, insertion of encrypted nodes, browsing of mindmaps, configuration of the tool, nodes formatting, searching for a node by label. It is integrated with a web browser. The possible import formats are: Branch; Linked branch: with or without root; Explorer favorites; Folder structure. The export formats supported are: HTML; XHTML; Open Office Document; PDF; JPEG; PNG; SVG. The arcs connecting nodes cannot be attached a label. The file format that is used for loading a map in the tool is .mm, which is XML format.</td>
<td>The tool is Open Source and its functionality can be extended. It is developed in Java. It is easy to use and offers a large number of functionalities. It is very comfortable for visual organization of different kinds of resources. It is related with the local file system and is integrated with a browser in order to allow browsing of web resources. If the files are a known Windows format (doc, ppt), the tool calls the suitable programs for browsing the local files. The other big advantage is the great variety of export formats. For the needs of the LTfLL Project the export format should be XML format. The tool supplies export to XHTML format which is an XML format and can be opened in a web browser. One disadvantage of the tool is that arcs cannot be labeled. Only hierarchies are possible to build that way. For the LTfLL project we aim to support more expressive visual model.</td>
</tr>
<tr>
<td>VUE (Visual Understanding Environment)</td>
<td>VUE is a 2D drawing application specialized for concept/content maps creation and presentation. It supports the following features: linking of resources to concept map's</td>
<td>The tool is Open Source and its functionality can be extended. It is developed in Java. It is very interactive and can be used for addition of functionalities that</td>
</tr>
</tbody>
</table>
D6.1 Social and informal learning support design

| License: Educational Community License | nodes (doc, ppt, html files, which can be opened or remote); browsing of linked to nodes resources; search for resources/nodes by specified keywords, categories, labels; multiple resources linking to a node; creation of presentation pathways - linear and non linear (very useful in case of big maps); playback a created pathway presentation; map display tools: zooming tool, full screen view, outline view, split screen view of multiple maps, pruning (hiding, showing branches); keywords and categories definition for nodes; search on keywords and categories in conjunctive and disjunctive mode; ontologies loading, attachment of ontological concepts to nodes; merging maps in voting or weight fashion. The export formats are: VUE, VUE package, html, pdf, jpg, png, IMS resource list - xml, zip (map only). The import format is .vue, which is an XML format. Created maps can be published: exporting, storing and sharing of VUE data in digital repositories. | cover our requirements. The features that are very useful are the keywords definition and search, connectivity with local and remote resources. We have contacted the development team to ask for a permission to extend the code with the following features: ontology presentation; documents visualization; lexicon translation; contexts definition and visualization. |
| SmartDraw Programming lang: C License: Proprietary License | Features: - Assisted Flowcharting: building flowcharts by clicking simple commands. - Creation of Business Graphics: SmartDraw includes SmartTemplates for charts and diagrams, including: organization charts, project charts, decision trees, cause-and-effect diagrams, genograms. - PowerPoint Integration with Auto Animation: SmartDraw includes a set of controls for optimizing graphics for PowerPoint. - Graphic Design - Charting: creation of charts and graphs without having to make a spreadsheet first - Image Charts: usage of pictures and images to display information instead of standard bars, lines, and circles - Interactive Maps: capturing live data from the Internet to incorporate roads, regions, counties, zip codes, countries and satellite images from across the The tool is proprietary. It can't be extended with any functionality. |
### Thinkgraph
**Programming lang:** Visual Basic 6  
**License:** Free Software Terms

ThinkGraph is a mix between a 2D Drawing application and a Concept Map editor. The tool supports: 2D Drawing (creation of 2D basic shapes: Line, Rectangle, Ellipse, Polygon, Text, Bitmap Image (PNG or JPEG)); formatting of text shapes: edition of font size, font color and style (Bold, Italic, Underlined); rotation of shapes; bookmarking of web resources. The concept maps can be exported to bitmap (JPEG or BMP), image Map (Bitmap with clickable areas), xml format, svg format. The import formats are xml and svg.

The tool is Open Source and its functionality can be extended. It is developed in Visual Basic. It is not very interactive for use. The maps have a hierarchy structure. The nodes can be associated with links and bookmarks. Links to local documents are possible. The documents can be opened by the tool by following the link to the document. The tool offers a similar functionality as the freemind tool but the interface is not so interactive and easy to use and configure.

### VYM (View Your Mind)
**Programming lang:** C++  
**License:** Proprietary License

VYM (View Your Mind) is a tool to generate and manipulate concept/mind maps. The tool supports the following functions: 2D Drawing - creation of 2D basic shapes: Line, Rectangle, Ellipse, Polygon, Text, Bitmap Image (PNG or JPEG); formatting of text shapes: edition of Font Size, Font Color and Style (Bold, Italic, Underlined); addition of icons, bookmarks to nodes. The import formats are map directory and xml. The export formats are Bitmap (JPEG, BMP,PBM,PGM,PNG,PPM,XBM,XPM), xml and XHTML formats.

The tool showed bugs even on basic functionality like association of a node with link.

### Kdissert
**Programming lang:** Python  
**License:** QPL (KDE4 libraries are not in use yet)

Semantik (previously Kdissert) is a mindmapping-like tool to help users to create: presentations, dissertations, theses, reports in an interactive and understandable view. The tool is available for Windows, Linux and other free operating systems.

<table>
<thead>
<tr>
<th>Tool Name</th>
<th>Description</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHOE Annotation Editor</td>
<td>The Knowledge Annotator is a Java program that allows users to mark-up web pages with SHOE knowledge. The Annotator is available as an applet or a stand-alone Java application. Documents are annotated with SHOE Ontologies (<a href="http://www.cs.umd.edu/projects/pl">http://www.cs.umd.edu/projects/pl</a>)</td>
<td>The tool uses and manipulates resources that are too restricted in their format: SHOE Ontology format, which is an extension of the HTML Syntax. The Ontology format of interest for the LTFLL project is OWL-DL, because of its expressive power, reasoning, existence of tools manipulating</td>
</tr>
</tbody>
</table>
### Annotea Protocol

**License:** OS

**Annotea** is a framework that supplies a Java library for communicating with the Annotea server. Annotations, according to Annotea, mean comments, notes, explanations, or other types of external remarks that can be attached to Web document. When the user gets the document he can also load the annotations attached to it from a selected annotation server or several servers and see what his peer group thinks. The annotations are stored in annotation servers as metadata and presented to the user by a client capable of understanding this metadata and capable of interacting with an annotation server via HTTP service protocol. The implementations of Annotea are divided into "client" and "server" ones.

### Annozilla

**Programming lang:** javascript

**Annozilla** is designed to view and create annotations associated with a web page, as defined by the W3C Annotea project. The idea is to store annotations as RDF on a server, using XPointer-like constructs. The intention of Annozilla is to use Mozilla's native facilities to manipulate annotation data - built-in RDF handling to parse the annotations and nsIXmlHttpRequest to submit data when creating annotations. Annozilla is an add-on to the Mozilla Firefox browser. The functionality, that should be supplied, is: loading of annotation; creation of annotations; posting

The tool is useful for organizing Web resources but doesn't satisfy the basic need of the LTfLL project: support of XML documents annotation, annotating with instances defined for a OWL-DL Domain Ontology concepts, loading of additional language and knowledge resources: Ontologies and Lexicons. During our investigation, we have experienced problems in connecting to an annotation server and local storage of annotations.
## D6.1 Social and informal learning support design

| **Amaya**<br>**Programming lang:** C | **Amaya** is a Web editor. Browsing features are integrated with the editing and remote access features in the environment. Amaya supports annotation of HTML, XML, XHTML, MathML, and SVG documents. The application is based on Resource Description Framework (RDF), XLink, and XPointer. | The tool is more HTML editor than an annotation tool. The resources that can be annotated are restricted in format: HTML, SVG, XHTML documents. XML documents can be loaded but their content cannot be annotated. The annotations (metadata) are fixed type: Advice, Change, Comment, Example, Explanation, Question, SeeAlso. There is no possibility to load ontologies and annotate XML document content with instances of ontology concepts. On testing we experienced the following problems: the posting of annotations to an annotation server was not successful; local storage of annotations was not successful; annotations content types are just the mentioned categories. |
| **Annotea Ubimarks in Mozilla/Firefox** | **Annotea Ubimarks** is part of Annotea social bookmarks and topics work in Mozilla. It makes users familiar with the common bookmark user interface metaphor to create metadata for Semantic Web, to share and combine bookmark data and bookmark categories, or topics from several locations or with other metadata. Topics in Annotea can be simple tags or they can form hierarchies. | The tool is useful for organizing Web resources but does not satisfy the basic requirements of the LTfLL project: - annotation of parts of the document - support of XML documents annotation - annotating with instances defined for OWL-DL Domain Ontology concepts - loading of additional language and knowledge resources: Ontologies and Lexicons |
| **Server Implementations** | **Zope Annotation Server Programming lang:** Python<br>**License:** Zope Public License | This product provides an annotation server for Zope, implementing the W3C’s Annotea protocol which allows clients to add annotations to web pages without modifying the pages’ source. Zope is an open source web application server. It features a transactional object database which can store not only content and custom data, but also dynamic HTML templates, scripts, a search engine, and relational database (RDBMS) connections and code. Zope also features an integrated security model, allowing to turn control over parts of a web site to other organizations or individuals. The transactional model happens automatically and applies to Zope’s object database and many relational database connectors. |
| **W3C Annotea Server’** | **Annotea** is a framework that supplies a Java library for communicating with the Annotea server. Annotation clients need to follow the Annotea protocols and understand the annotation schema. The user can install his... |
### CLaRK

**Programming lang:** Java  
**License:** OS

CLaRK is an XML-based software system for corpora development implemented in Java. The main aim behind the design of the system is the minimization of human intervention during the creation of language resources. It incorporates several technologies: XML technology; Unicode; Regular Cascaded Grammars; Constraints over XML Documents. The core of CLaRK is an Unicode XML Editor, which is the main interface to the system. Besides the XML language itself, it is implemented an XPath engine for navigation in documents and an XSLT processor for transformation of XML documents. For multilingual processing tasks, CLaRK is based on an Unicode encoding of the information inside the system.

The basic mechanism of CLaRK for linguistic processing of text corpora is the cascaded regular grammar processor. The main challenge to the grammars in question is how to apply them on XML encoding of the linguistic information. The system offers a solution using an XPath language for constructing the input word to the grammar and an XML encoding of the categories of the recognized words.

The insufficient expressive power of DTD for constraining documents’ content is compensated by constraint tools which can work in insertion and validation mode. They are more powerful than XML Schema restrictions.

---

The tool is implemented with the intent to make the language resources (corpus of documents, lexicons, ontologies) manipulation automatic or semi-automatic. The technologies used and integrated are: XML technologies, XPath technologies. The tool is used for annotation with ontological data of corpora. The tool (basically Grammars, Constraints, Concordance) is used in the context of the LT4el (http://www.lt4el.eu/) and the AsIsKnown (http://www.asisknown.org/) projects. NLP tools can be connected in a cascaded manner (language pipes) in order to process language resources consecutive.
Appendix B

This appendix contains an overview of the most relevant social media applications available and the communities which have grown around them. The classification was inspired by Marlow et al. (2006), whose classification criteria are used as the upper 7 criteria from the table below. "Social Connectivity" was replaced by "Explicit Social Connectivity", to distinguish it from implication connectivity, i.e. "Derived Relations" that are derived by the system, such as a community of users who have all used a certain tag.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tagging Features</td>
<td>free-text</td>
<td>choose one category</td>
<td>free-text</td>
</tr>
<tr>
<td>Tagging Rights</td>
<td>free-for-all</td>
<td>self-tagging</td>
<td>free-for-all</td>
</tr>
<tr>
<td>Tagging Support</td>
<td>suggested tagging: &quot;recommended&quot; and &quot;popular&quot; tags for this resource are shown, as well as &quot;all my tags&quot;; you can insert a suggested tag by clicking on it</td>
<td>viewable</td>
<td>suggested</td>
</tr>
<tr>
<td>Aggregation Model</td>
<td>bag</td>
<td>n/a (self-tagging)</td>
<td>bag</td>
</tr>
<tr>
<td>Object Type</td>
<td>bookmarks</td>
<td>bookmarks</td>
<td>bookmarks</td>
</tr>
<tr>
<td>Source of Material</td>
<td>web link</td>
<td>web link</td>
<td>web link</td>
</tr>
<tr>
<td>Resource Connectivity</td>
<td>none</td>
<td>groups</td>
<td>none</td>
</tr>
<tr>
<td>Explicit Social Connectivity</td>
<td>• following (people in network)</td>
<td>• mutual friends – friends agreed by both sides</td>
<td>• friends</td>
</tr>
<tr>
<td></td>
<td>• followers (fans)</td>
<td>• fan – user added by the current user to his/her network</td>
<td>• groups</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Derived Relations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• top-11 related tags for a tag</td>
<td>• people digging the same story</td>
<td>• top tags for a user</td>
</tr>
<tr>
<td></td>
<td>• all users and tags for a certain bookmark</td>
<td>• people who comment on the same resources</td>
<td>• tag community</td>
</tr>
<tr>
<td></td>
<td>• all tags for a user</td>
<td></td>
<td>• site community (users who have saved bookmarks of a certain site, e.g.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• people who recently bookmarked the same websites</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• recommended bookmarks</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• friends, groups, tags, bookmarks that two people have in common</td>
</tr>
<tr>
<td>Private Interaction Features</td>
<td>recommend a bookmark to a person</td>
<td>messaging</td>
<td>messaging</td>
</tr>
<tr>
<td>Public Interaction Features</td>
<td>comments on resources</td>
<td>blogging</td>
<td>user's pinboard</td>
</tr>
<tr>
<td></td>
<td>• public personal profiles (only user's name, mail address and homepage)</td>
<td>comments on resources</td>
<td>public personal profiles</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• user's pinboard</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• public personal profiles</td>
<td></td>
</tr>
<tr>
<td>Estimated number of users</td>
<td>1,700,000 visitors/month</td>
<td>236 million/year (en.wikipedia.org)</td>
<td>?</td>
</tr>
<tr>
<td>Remarks</td>
<td></td>
<td></td>
<td>an enormous amount of options, such as highlighting, sticky notes, import/export interface to other bookmarking websites, subscription to bookmarks with a certain tag</td>
</tr>
<tr>
<td>----------------------------</td>
<td>----------------------</td>
<td>---------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td><strong>Tagging Features</strong></td>
<td>free-text</td>
<td>free-text</td>
<td>free-text</td>
</tr>
<tr>
<td><strong>Tagging Rights</strong></td>
<td>free-for-all</td>
<td>self-tagging</td>
<td>self-tagging</td>
</tr>
<tr>
<td><strong>Tagging Support</strong></td>
<td>blind (default), suggestions from top-tags shown after clicking</td>
<td>blind</td>
<td>blind</td>
</tr>
<tr>
<td><strong>Aggregation Model</strong></td>
<td>Bag</td>
<td>n/a (self-tagging)</td>
<td>n/a</td>
</tr>
<tr>
<td><strong>Object Type</strong></td>
<td>book metadata</td>
<td>video</td>
<td>content/comments</td>
</tr>
<tr>
<td><strong>Source of Material</strong></td>
<td>system (catalog) and user-contributed</td>
<td>user-contributed</td>
<td>user-contributed</td>
</tr>
<tr>
<td><strong>Resource Connectivity</strong></td>
<td>None</td>
<td>resources are linked according to domain</td>
<td>links between blogs - can be obtained from technorati.com who computes an authority for each blog</td>
</tr>
<tr>
<td><strong>Explicit Social Connectivity</strong></td>
<td>• friends</td>
<td>• friends</td>
<td>• blogroll - list of blogs that are read by the owner of the blog</td>
</tr>
<tr>
<td></td>
<td>• group members</td>
<td>• subscriber – a user subscribes to another user as some kind of fan</td>
<td>• comments</td>
</tr>
<tr>
<td><strong>Derived Relations</strong></td>
<td>• people who saved the same book</td>
<td>people who comment on the same resources</td>
<td>• links between people and blogs</td>
</tr>
<tr>
<td></td>
<td>• top tags for a book</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Private Interaction Features</strong></td>
<td>None</td>
<td>mail messaging</td>
<td>mail messaging (optional)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Public Interaction Features

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>forum (discussion threads about a book)</td>
<td>comments on resources</td>
<td>blogging</td>
</tr>
<tr>
<td></td>
<td>user's pinboard</td>
<td>user's pinboard</td>
<td>comments on resources</td>
</tr>
<tr>
<td></td>
<td>public personal profiles</td>
<td>public personal profiles</td>
<td>user's pinboard</td>
</tr>
<tr>
<td></td>
<td>rating/review of content</td>
<td>rating of content</td>
<td>public personal profiles</td>
</tr>
</tbody>
</table>

### Estimated number of users

|                          | ?                     | 24,000,000/month | ?                             |

### Remarks

- can be on a public platform like wordpress.com or blogger.com or can be hosted on a personal server.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tagging Features</td>
<td>n/a</td>
<td>free-text tags + content type category (3 fixed values)</td>
<td>choose one category</td>
</tr>
<tr>
<td>Tagging Rights</td>
<td>n/a</td>
<td>permission-based viewing + tagging (you can tag what you can view)</td>
<td>self-tagging</td>
</tr>
<tr>
<td>Tagging Support</td>
<td>n/a</td>
<td>viewable</td>
<td>viewable (the category is chosen from a taxonomy)</td>
</tr>
<tr>
<td>Aggregation Model</td>
<td>n/a</td>
<td>set</td>
<td>n/a (only one category)</td>
</tr>
<tr>
<td>Object Type</td>
<td>short messages (max 160 characters)</td>
<td>images</td>
<td>questions</td>
</tr>
<tr>
<td>Source of Material</td>
<td>user-contributed</td>
<td>user-contributed</td>
<td>user-contributed</td>
</tr>
<tr>
<td>Resource Connectivity</td>
<td>None</td>
<td>groups (images belong to a set)</td>
<td>groups (questions are in a category)</td>
</tr>
<tr>
<td>Explicit Social Connectivity</td>
<td>• followers</td>
<td>• friends</td>
<td>friends</td>
</tr>
<tr>
<td></td>
<td>• following</td>
<td>• groups</td>
<td></td>
</tr>
<tr>
<td>Derived Relations</td>
<td>None</td>
<td>• images for a tag</td>
<td>tag community / related tags</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• related tags</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• all tags for an image</td>
<td></td>
</tr>
<tr>
<td>Private Interaction Features</td>
<td>Messaging</td>
<td>mail messaging</td>
<td>• mail messaging</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• instant messaging</td>
</tr>
<tr>
<td>Public Interaction Features</td>
<td>• blogging</td>
<td>comments on resources</td>
<td>• answers on questions</td>
</tr>
<tr>
<td></td>
<td>• public personal profiles</td>
<td></td>
<td>• rating of answers</td>
</tr>
<tr>
<td>Estimated number of users</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Remarks</td>
<td>probably most similar to chats due to the shortness of messages</td>
<td></td>
<td>The &quot;best answer&quot; is the answer that is voted as the answer that best solves the question. An automatic user profile keeps track of the number of posted questions, answers and &quot;best answers&quot;.</td>
</tr>
</tbody>
</table>
Appendix C

Several projects are emerging at the European level that can produce interesting results which can help us reach the objectives of our Workpackage. We have an ongoing collaboration with the projects mentioned below in order to exchange results and information.

**Aposdle**
http://www.aposdle.tugraz.at/
Duration: 03/2006 – 03/2010
Main features:
- Lifelong learning: Learn within the context of your immediate work and within your current work environment.
- Distinction with respect to traditional eLearning systems: integrated technological support for all three roles a knowledge worker fills at the workplace: the role of the worker, the role of the learner, and the role of the expert.
- System that proposes "knowledge artifacts" and brings together users within an organization with three roles: worker, expert and knowledge owner, so that they can collaborate. The system supports "collaboration sessions".

Resulting resources:
- Tools that locate suitable knowledge sources.
- Tool for annotation of collaboration session (i.e. chat transcript) with fixed-vocabulary keywords.
- Domain Modeling Tool, for concept annotation of documents, as a Protégé plugin.
- Integrated Semantic Knowledge Structure (compare to LTfLL CSF) and Associative Network (= ontology)

**iCamp**
http://www.icamp.eu
Duration: 10/2005 – 10/2008
- The project pursues the idea of gathering people (learners, facilitators, peers, etc.) into one common virtual learning environment. This virtual environment does not consist of a single software system, but is composed of various interoperable tools and platforms.

**MELT**
http://info.melt-project.eu/ww/en/pub/melt_project/welcome.htm
Duration: 10/2006 – 12/2008
- Content enrichment (adding metadata), including social tagging of learning resources; also automatic metadata generation (different kinds of metadata than in LT4eL)

Resulting resources:
- Over 37,000 learning resources and 97,000 learning assets with Creative Commons licenses
- Tools and procedures for social tagging of learning material
BONy
http://www.bonynetwork.eu/
Duration: 01/2008 - 12/2009
Main features:
- LMS
- Multilingual access
- Ontological approach with semantic web technology
- Social network
- Users can play double role: student/teacher

TenCompetence
http://www.tencompetence.org/
- Publications: http://dspace.ou.nl/handle/1820/497/browse-date

The project will (among other things) develop software for the effective support of users who create, store, use and exchange knowledge resources, learning activities, units of learning and competence development programmes within a learning network

Resulting resources:
- Personal Competence Manager (http://www.tencompetence.org/PCM/download.html)
  - Use competence profiles, which have to be kept up-to-date. Users have to indicate their own competence with levels.
- TENCompetenceTube (http://www.calt.insead.edu/eis/tencompetencetube/)
  - Connecting people to people (social network dimension)
  - Connecting people to Competence-related Knowledge
  - Use of videos and visualization of social networks to achieve this