Ontology Maturing: a Collaborative Web 2.0 Approach to Ontology Engineering

Simone Braun
FZI Research Center for Information Technologies
Karlsruhe, Germany
braun@fzi.de

Andreas Schmidt
FZI Research Center for Information Technologies
Karlsruhe, Germany
aschmidt@fzi.de

Andreas Walter
FZI Research Center for Information Technologies
Karlsruhe, Germany
awalter@fzi.de

ABSTRACT

Most of the current methodologies for building ontologies rely on specialized knowledge engineers. This is in contrast to real-world settings, where the need for maintenance of domain specific ontologies emerges in the daily work of users. But in order to allow for participatory ontology engineering, we need to have a more realistic conceptual model of how ontologies develop in the real world. We introduce the ontology maturing processes which is based on the insight that ontology engineering is a collaborative informal learning process and for which we analyze characteristic evolution steps and triggers that have users engage in ontology engineering within their everyday work processes. This model integrates tagging and folksonomies with formal ontologies and shows maturing pathways between them. As implementations of this model, we present two case studies and the corresponding tools. The first is about image-based ontology engineering (introducing so-called imagenotions), the second about ontology-enabled social bookmarking (SOBOLEO). Both of them are inspired by lightweight Web 2.0 approaches and allow for realtime collaboration.

Categories and Subject Descriptors

H.1.m [**Miscellaneous**]; H.5.3 [**Group and Organization Interfaces**]; I.2.4 [**Knowledge Representation Formalisms and Methods**]: Semantic networks; K.4.3 [**Organizational Impacts**]: Computer-supported collaborative work

General Terms

Design, Human Factors, Theory

Keywords

Ontology engineering, ontology maturing, Web 2.0, tagging, social software, work integration, collaboration

1. INTRODUCTION

Within state-of-the-art semantic approaches, ontologies have emerged as the key to enable more advanced technological support for end users and their work processes, which particularly applies to knowledge work. However, current research and development concentrates more on what we can do as soon as we have ontologies—rather than having a closer look at the processes of creating and especially maintaining such domain-specific ontologies. In real-world settings these issues are crucial to fulfill the users' needs and currently insufficiently dealt with.

It is usually acknowledged that ontologies are shared understandings of a particular domain that have to be constructed within social processes among the stakeholders.

However, current methods and tools do not empower the users to actually carry out these negotiation processes embedded in their work.

This is mainly due to lopsided and partially naive perspectives on ontology engineering. One perspective requires ontology engineering experts who moderate the ontology creation processes and on whom users will depend indefinitely. In the other perspective, it is assumed that users become ontology modeling experts right away. Presupposing an ontology engineering expert, we have to face two problems: First, involving ontology engineering specialists is expensive and second, ontologies produced by modeling experts instead of domain experts can contain errors caused by the insufficient domain knowledge and experience of knowledge engineers [2]. Leaving the complete, challenging and complex task of ontology modeling to the domain experts on the other hand is usually not an option either! Often users are not motivated to invest the effort because they are concerned with their work processes and regard ontology modeling in its traditional form as an overhead. The main reason therefore is the fact that the time lag between the emergence of concepts and their inclusion in ontologies is far too big for ontologies to be useful [14]: concepts are already becoming obsolete by the time they are entering the ontology.

Indeed, users are almost constantly constructing and negotiating shared meaning in collaboration with others by augmenting and evolving a community vocabulary. The main challenge is then how to leverage this implicit and informal ontology building for the explicit formal models needed for semantic approaches. What we actually need in order to cope with this challenge are two things: First, on the conceptual level: a more realistic and work-integrated view of how ontologies actually are or can be created and second, on the technical level: tools supporting interweaving working and ontology engineering activities and the associated social negotiation process.

This paper is organized as follows: Section 2 discusses related work. In section 3, we present our notion of ontology maturing as a conceptual model and analyze the motivation for users, to participate on such collaborative maturing processes. We present two case studies in section 4 and 5—

WWW2007, May 8–12, 2007, Banff, Canada.
.

they show how this conceptual model in combination with lightweight and collaborative Web 2.0 tools enables and fosters the maturing process. Section 6 concludes.

2. RELATED WORK

There are several proposed methodologies for the process of ontology engineering that require ontology engineers as moderators. These methodologies define how to support the ontology lifecycle from development, via evaluation and maintenance, to further evolution. An overview is given in [7] and [10]. They do not allow for the work-embedded and collaborative engineering of ontologies

A more "human-centered approach" is taken by [17] with the Human-Centered Ontology Engineering Methodology (HCOME). Kotis et al. [17] view ontology development as a dynamic process and focus particularly on ontology evolution. They assume a decentralized engineering model where everyone first formalizes her own ontology and shares it in a further step within the community. There, the individual ontologies are merged or further developed. However, findings in [1] (based on action theory) suggest that collaboration plays a more important role *before* we have formalized (individual) ontologies. So we think that the HCOME methodology can benefit from incorporating the notion of different maturity levels. This methodology does not support our goal of embedding ontology engineering of ontologies in work processes.

Editing tools like Protégé [22] or KAON OIModeler [19] are commonly used for ontology building ([21] and [6] discuss them in detail). Most existing ontology editors are standalone desktop applications. In addition, they consider ontology construction as an isolated and detached task where ontology engineering experts explicitly sit together with knowledge workers to model the ontologies. They do not provide a collaborative environment or support collaboration only in a restricted way (except for KAON). Like the above mentioned methodologies, these tools are not geared towards knowledge workers and their work processes.

Wiki systems consider the aspects of collaboration and can support the early phases of ontology construction. Semantic wiki systems[1] try to extend the traditional wikis with semantic web technologies (e.g. Semantic MediaWiki [28], OntoWiki [15] or IkeWiki [23]). These systems help users in creating definitions, e.g. beginning with informal texts. Because of discussion pages and versioning for each article they are suitable for complex coordination and consolidation processes, but they are usually more time-consuming and lack work process integration.

Collaborative tagging systems form a third application group that is getting more and more popular on the web. Their lightweight approach allows users to easily assign keywords to various contents. The social bookmarking systems BibSonomy ([16]) and Del.icio.us [5] further allow grouping of tags underneath a super-tag. However, they do not define a clear concept structure and this functionality is very restricted. For instance these applications do not display the tree structure formed by these relations and the structure on top of tags is not shared.

There were first attempts to analyze occurrence patterns and in particular the usage of tags within collaborative tagging systems using the example of Del.icio.us [5] and Flickr [8]. Golder & Huberman [9] observed that tags used for an individual resource stabilize over time. Marlow et al. [20] showed that the tag vocabularies of socially connected users have a bigger overlap than those of randomly selected users. Both might be traced back to issues of imitation and an upcoming shared knowledge. Sen et al. further investigated the community impact on personal vocabulary and tagging behavior [25]. They explored different forms of selecting tags used in the community. They observed that users apply ("borrow") an already used tag more often in case tags of community members compared to when no tags are displayed. Our model in the next chapter considers these observations as a change from ideas to consolidation in communities.

[1]See website of the SemanticWiki Interest Group for updated list of existing semantic wiki systems [12]

3. THE ONTOLOGY MATURING PROCESS

In the following, we will present the steps of our model and the motivation and triggers for user to participate on the ontology maturing progress.

3.1 Observations about ontologies

At first, we will collect some important observations about ontologies, based on our experiences, that help in developing a more appropriate model for building ontologies:

- **Ontology building as a learning process.** Key is the insight that building ontologies is not just about eliciting knowledge and formalizing it according to a particular formalism. Rather this construction process itself is a learning process in which the involved individuals deepen their understanding of the real world and of an (appropriate) vocabulary to describe it. This is especially true if we consider emerging ideas: it is simply not possible to integrate these into an ontology straight away.

- **Formality and complexity of use as barrier.** Related to that learning process, we should also consider the lessons learnt from the success of Web 2.0 approaches. Web 2.0 approaches empower the individual to take part in community activities by lowering the barriers: informal, lightweight, easy-to-use, and easy-to-understand. Tagging as an organizing paradigm has replaced folder hierarchies. As a conclusion, we should acknowledge that the degree of formality poses a barrier. And that the compelling simplicity of Web 2.0 applications is part of their success model.

- **Continuous evolution in work processes.** Ontology building is usually not supposed to be a one-time activity of an expert committee, but rather a sustainable process of continuous evolution. But if we cannot assume that everything is on the same formality level, we need to acknowledge that in a living ontology there are most likely concepts not clearly defined yet. They underlie a process of continuous evolution where ideas and understanding emerge implicitly in daily work and mature only gradually through the interaction with others to explicit formal and shared conceptualizations.

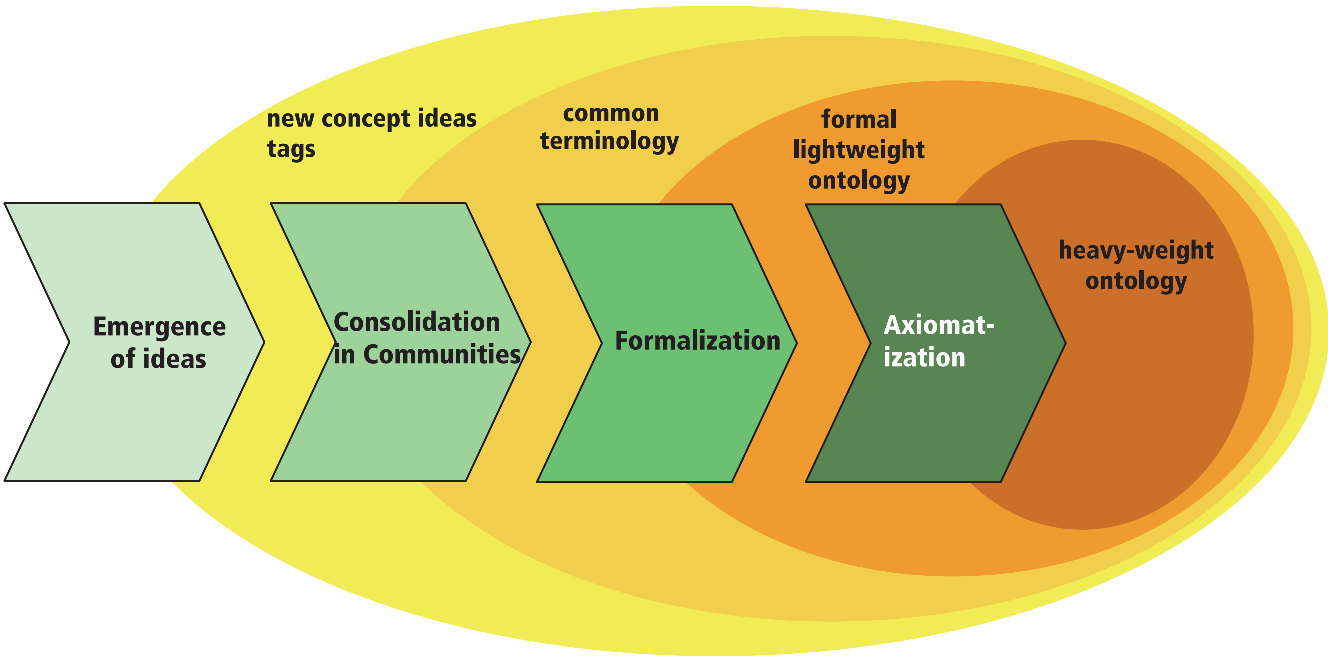

Figure 1: Ontology Maturing Process

3.2 The Model of Ontology Maturing

These observations are similar to those made in [24] about how new ideas develop in the context of knowledge management and e-learning to become reusable training material. This development process was described with the metaphor of *maturing* and structured into five phases as the so-called knowledge maturing process. This process is viewed as a macro model for interconnected individual learning processes. Based on this process model, we have come up with a similar phase model that identifies characteristical maturing transitions in collaboratively developing a shared ontology (see Figure 1):

1. **Emergence of ideas.** In this initial transition, new concept ideas are introduced which are rather ad hoc and not well-defined. They are personal “utterances” which are informally communicated and technically typically represented by tags. For instance, while annotating or seeking for resources, we recognize that the tag we want to use does not exist or is misspelled. Accordingly, we introduce a new tag or correct the existing one without further reflecting.

2. **Consolidation in Communities.** Through reuse and adaption of concept symbols (tags) of others, a shared vocabulary emerges within a community. When comparing currently envisioned tags with previously used ones or with tags from other people assigned to the same resource, we discover similarities and differences that allow for creating concepts from tags. For instance, we realize that we can improve our search by using a synonymous term. We establish a link between these terms in our understanding and thus can merge synonyms into concepts. Moreover, preferred labels also often develop in the same way [9]. It crystallizes a common terminology still without formal semantics.

3. **Formalization.** Within the third phase, concepts are organized into relations - both taxonomical (hierarchical) and ad hoc relations. For instance, we want to receive recommendations from the system and we become aware of different abstraction levels (e.g. broader and narrower) that we need to have in order to find e.g. spaghetti and bigoli (thicker spaghetti) as noodles and vice versa. The results are lightweight ontologies that rely primarily on inferencing based on subconcept relations.

4. **Axiomatization.** The last phase of our model captures more domain semantics by adding background knowledge for improving inferencing processes, e.g. for query answering. This step requires a high level of competence in logical formalism so that this can usually only be carried out by domain experts.

3.3 Motivation and Triggers for Maturing Activities

This model as such only describes characteristic degrees of formality of ontology elements from a user-driven point of view. But it does not explain yet, how this integrates with actual work processes.

When and how do users engage in maturing activities? What triggers the maturing of concepts and other ontology elements? And why do we believe that users are motivated to participate in this collaborative maturing effort?

Starting point for our analysis is the observation that ontologies are mainly used and deliver their main benefit for information seeking or distribution activities. This includes users wanting to:

- retrieve appropriate content
- promote their content so that it is found by others

- enhance their work performance (getting answers instead of searching)
- be empowered by being able to state their opinions

This coincides with motivations for tagging resources in [20]: Future retrieval, Contribution and Sharing, Attract Attention, Play and Competition, Self Presentation, and Opinion Expression.

The embedding of ontology maturing into users work processes also requires to have a look at the micro level of concrete triggers initiating these activities. These triggers are important to consider because they are the technical integration points where we need to make the transition between working and ontology maturing possible.

If we consider the retrieval of appropriate content, we can identify following typical triggers which are closely connected to semantic shortcomings of tagging systems (cf. [9], [13]):

- **(Mis-)Spelling.** The most obvious problem is that tags were simply misspelled or that there are typographical differences, for instance because of occurring plurals, abbreviations or compound words. Even inexperienced users tend to correct such issues if possible.
- **Synonymy.** Documents are not found with the first keyword, but later under a synonym. Similar is the issue for the author of a resource trying to promote its usage: she has to use many synonyms in order to ensure that other users will find it later on. This typically leads users to abstract from keywords to concepts with several synonyms (possibly also in multiple languages).
- **Multilingualism.** Tags are always in one language. This requires that content owners, especially in Europe with many different languages, have to describe resources with many tags in different languages.
- **Homonymy.** A tag can have different meanings, so called homonyms. This leads to search results with low precision, as all resources that are relevant to these different meanings are annotated with the same tag. Fixing this issue is more demanding as it requires to differentiate existing tags into different concepts. Its effects can only be realized if resources are reindexed, which often is not manageable.
- **Missing concepts.** During annotation activities, users discover that a topic they want to describe is not yet covered. This leads to additions to the ontology.
- **Mismatch of abstraction level.** A typical search problem in complex domains is that search terms are specified either too broad or too narrow. This problem, also known as the "basic level phenomenon" [27], can be traced back to different intentions and expertise levels of the users. For instance, one user tags a resource on the basic level with "spaghetti", another with "noodles" and a third differentiates "spaghetti" from "bigoli" (thicker spaghetti) and "vermicelli" (thinner spaghetti). Thus, later success in finding leads to the adding of hierarchical relationships.
- **Missing guidance to related information.** Especially the orientation phase of information seeking processes (like [18] described) requires guidance to find related fields. As soon as implicit links are discovered, they can be added to the ontology. An increasing number of links (at the latest) leads to defining new properties.

3.4 Tool support

In the following two sections, we present two approaches that support the collaborative maturing process of ontologies based on our introduced model and the relevant triggers. Both provide tool support in work integrated scenarios.

The first case study is rooted in the domain of semantic image retrieval. Here, the focus lies on the *consolidation of image descriptions in communities* with the help of visualized concepts and relations. The case stuy introduces a new method called *imagenotions*. This method allows the collaborative maturing of unstructured tags to commonly accepted concepts.

The second approach also provides consolidation capabilities, but in particular concentrates on how to support the formalization based on consolidated concepts. The key idea here is to couple social bookmarking approaches with a lightweight ontology editor. Its application domain is informal learning and knowledge management support in interdisciplinary application-oriented research.

4. IMAGINATION

The IMAGINATION EU project[2] provides image-based navigation for digital cultural and scientific resources. Users can click on parts of an image to find other interesting images to a given context. A click on an image part automatically generates a semantic query. This feature requires semantic metadata. In IMAGINATION, semantic metadata are automatically generated by combining text-mining, object recognition and object identification algorithms exploiting domain ontologies. The metadata can be validated, corrected and extended manually if needed.

4.1 Requirements

Based on our experience in implementing commercial image search systems[3], current metadata about the content of images is largely based on the unstructured and non-semantic tagging paradigm (even before the popularity of Web 2.0)—often called also keyword or label and embedded into IPTC attributes [4]. This tag-based annotation of images has many of the shortcomings mentioned as triggers for entering ontology engineering activities in the previous section. Previous attempts to solve the problem with the help of thesauri have largely failed because thesauri were not accepted by the users and content providers due to their incompleteness and/or complexity.

These experiences can be directly fed into requirements for the use of ontologies in the area of image search and navigation: (1) the ontology (and therefore semantic annotations using it) must be comprehensible for the users and (2) an ontology must cover the image repository completely. The challenge is that image repositories can change rapidly and consequently the ontology must be adapted regularly.

[2] http://www.imagination-project.org

[3] See e.g. www.fotomarktplatz.de

4.2 Ontology Development Methodology Based on Imagenotions

Based on the ontology maturing process, we can develop a solution for these problems. In terms of the previously introduced ontology maturing process we concentrate on the first three steps from the emergence of ideas up to formalization. What distinguishes our methodology from the usual ontology development methodologies is the strong emphasis on collaborative ontology development in the consolidation-in-communities phase. This is motivated by the success of collaborative tagging in Web 2.0 projects.

The basis of our ontology formalism is a concept we call *imagenotions*. An imagenotion graphically represents a semantic notion through an image, or a set of images. In addition, similarly to many existing ontology formalism, it is also possible to associate tags with an imagenotion in different languages (such as English or German). For each language, one of these tags is selected as the main label of the imagenotion. The other tags are termed synonyms.

Instead of tags, images are annotated with imagenotions. It is easy to see the advantage: all the shortcomings of tagging approaches are solved using imagenotions—it is easy to find images using search terms in any language. In other words, we provide semantic search instead of full-text search.

In the terminology of classical ontologies, imagenotions are usually instances, but they may also correspond to concepts or relations. There are two major advantages of using imagenotions over the classical ontology constructs:

1. The distinctions between concepts, instances and relations are hard to understand for most users. In our mind, notions play the role of an instance, a concept or a relation, depending on the actual context. This fact is acknowledged by many ontology formalisms that allow metamodeling. Using imagenotions, users do not need to understand this somewhat artificial separation of notions.

2. Because imagenotions are associated with images, they are meaningful internationally as an image has the same meaning in different languages.

The goal of our methodology is to guide the process of creating an ontology of imagenotions. The main steps of this methodology is based on the ontology maturing process model:

1. *Emergence of Ideas.* In this step, new imagenotions are created. Already this step can become collaborative, as users can jointly collect the tags describing imagenotions, and select the most representative images for an imagenotion. Collaborative editing is especially useful in a multi-lingual environment where it cannot be expected that any individual user speaks all required languages.

2. *Consolidation in Communities.* Because it is so easy to create new imagenotions, it cannot be avoided that for the same semantic notion initially many imagenotions are created (synonyms, also in different languages) or that an imagenotion represents more than one semantic notion (homonyms). In this step, these problems should be solved by merging synonymous imagenotions, and by splitting imagenotions representing more than one notion.

3. *Formalization.* In this step, taxonomical ("is-a") and ad-hoc relations are specified among imagenotions.

After step 2, the quality of image search already increases significantly, as the problems with synonyms and homonyms do not appear anymore. Moreover, it is easy to see that all annotated images automatically benefit from the maturing imagenotions. E.g. adding a new tag to an imagenotions automatically allows users to find all images annotated with that imagenotion using the new tag. In addition, the outcome of step 3 also allows requests for related images based on the current context.

Imagenotions are useful to collaboratively build an ontology supporting manual annotation and semantic search. However, to fulfill the requirement of IMAGINATION for automatic annotation, a classical formal ontology is needed that can be exploited by text-mining and object identification algorithms. This last axiomatization step is not yet directly supported by our methodology, it is subject of future research. Nevertheless, it is easy to see that a conversion of an imagenotion ontology to a standard ontology formalism (e.g. OWL) is possible. The only missing information is whether an imagenotion should be modelled as a concept, instance or a relation in the target ontology formalism.

4.3 Tool Support

Currently we implement a web-based tool that allows the creation of new imagenotions and the editing of available imagenotions. This tool supports all three steps of our methodology. It can be easily invoked during semantic search or when uploading new images into an image repository: e.g. it is fully integrated into the user's workflow.

We now demonstrate some functionality of the tool in terms of the steps of our development methodology.

4.3.1 Step 1: Emergence of Ideas

Figure 2 shows an example for the emergence of ideas. Let us assume that a content owner has new images about elephants. The imagenotion "elephant" was so far not available. Therefore, she creates a new imagenotion, adds an image or part of an image that shows elephants and starts describing the new imagenotion with more details. She uses English as spoken language. As synonyms, she enters "elephantidae" and "tusker". Instead of tagging the new images that show elephants with these words, she can use the new imagenotion—she just pulls this imagenotion over the new images via drag and drop.

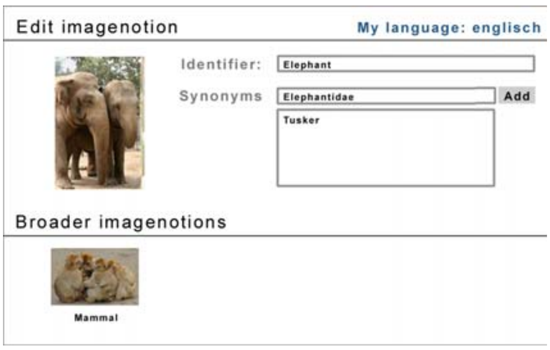

Figure 2: Editing an imagenotion with the NotionEditor tool

Figure 3: Spliting homononyms in different imagenotions

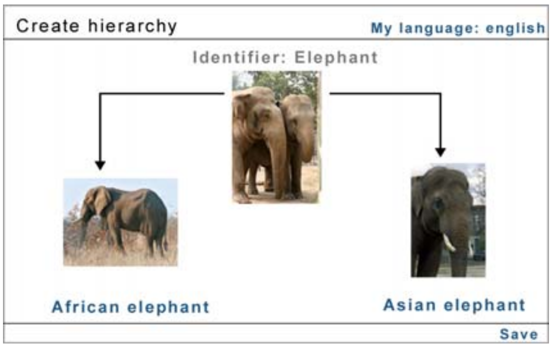

Figure 4: Creating hierarchies in imagenotions

Figure 5: Adding relations to imagenotions

Figure 6: Visual search refinement

The front end of the image search system also allows users to add missing synonyms to already defined imagenotions in order to achieve better search results. The label and synonyms of imagenotions are editable in different languages.

All images described with imagenotions automatically benefit from all added information. In other words, the community improves the quality of the image descriptions by improving the quality of the corresponding imagenotions.

4.3.2 *Step 2: Consolidation in Communities*

Consolidation solves the problem of homonyms and different languages. For instance, the word “tusker” is a homonym because (among others) it is also the brand of a Kenyan beer brewery (see Fig. 3). The first user noticing this homonym can solve the problem: she splits the imagenotion tusker in two new imagenotions—one with an image of an elephant for the meaning “elephant”, and another with the label of the Kenyan brewery for the meaning of the “Kenyan beer brand”.

In the case when someone detects two imagenotions that have the same meaning (possibly in different languages), she can merge them together.

4.3.3 *Step 3: Formalization*

Figure 4 shows an example for formalization—the creation of hierarchies. In this example two new imagenotions for two sorts of elephants are created: the “African elephant” and the “Asian elephant” imagenotions. The trigger for creating hierarchies may be the description of images in more details.

As elephants are mammals, one could also add broader imagenotions (the “is-a” relation): the user searches for the imagenotion “mammal” and adds it by drag and drop to the imagenotion of elephants to state that “every elephant is a mammal”. It is also possible to define other kinds of relations. For instance, in Figure 5 somone defined the relations “has baby” and used this relation between the imagenotions “Angelina Jolie” and her baby “Shiloh” as well as her son “Maddox Chivan”.

4.4 Application for Semantic Search

In addition to improving search results, imagenotions also allow a new principle of visual search refinement based on formal knowledge. Users start by typing some search terms. Our system proposes imagenotions containing these search terms as label or synonym. Furthermore, our system also proposes images and search refinements based on the current context. These proposals are visualized with imagenotions and determined using the relations from the available imagenotion ontology. Refining of a search request is possible by just clicking on the desired imagenotion.

Figure 6 shows such an example. The current search context is “images that contain the actors Angelina Jolie and Brad Pitt” and the user already selected a matching image. The system proposes two related groups of images: one for the film “Mr. and Mrs. Smith” and one for their common baby. These proposals are visualized using the corresponding image notions.

5. IM WISSENSNETZ – IN THE KNOWLEDGE WEB

Our second use case is taken from the German research project “Im Wissensnetz – Vernetzte Informationsprozesse in Forschungsverbünden"[4] ("*In the Knowledge Web – Networked Information Processes in Research Associations*"), which aims to support efficient interdisciplinary knowledge-added processes within e-Science. Research is likely to be the most knowledge-intensive environment. An empirical analysis of existing (cooperation-)processes, information and knowledge exchanges, and instruments for the preservation of knowledge accomplished in the application domain "rapid prototyping" revealed that scientific work is characterized by high variability, dynamics and unpredictability as well as by high significance of social interactions and communication.

5.1 Requirements

Linking people with individual expert knowledge and contents from various disciplines like plastics, ceramics, and mechanical engineering is one of the most important challenges. For instance, one major problem is searching and retrieving adequate contemporary resources. This process is very tedious. The users have to access many various data sources with different interfaces, but also the internet with common search engines like Google.

In the area of plastics and their market these high dynamics are particularly obvious. New materials or new forms of existing ones frequently enter the market; brand names and manufacturers are permanently changing and hardly trackable—attributes of a chemical substance retrievable using its brand name today, are very hard to find once it's sold under a different label. There is also no general up to date database which list manufacturer and brand names of currently available forms of plastics. Thus, the users rely on search engines like Google in order to find the wanted material of which only the old brand or chemical name is known. However, common search engines provide many irrelevant results because of their missing focus on the domain. For instance, when looking for "nylon" you receive lots of results for stockings.

At this point, using annotation and retrieval tools can help; e.g. when a colleague already found the new brand name of a product and tagged it with the old one you are looking for. With such tools being further semantically enriched with background knowledge and domain ontologies, it is possible to find out the search context and thus to extend or refine the search in order to reduce irrelevant results and to guide the user. However, this also requires that (1) the users collaboratively build up and maintain a shared understanding and terminology, (2) these activities are embedded into their information seeking activities and (3) this understanding is expressed formally enough to enable ontology-based query refinement.

5.2 Ontology Development Methodology Based on Semantic Social Bookmarking

In order to overcome these problems, we provide a solution that combines collaborative collecting and sharing of web resources (bookmarks) with collaborative development of a shared vocabulary. This vocabulary is used to organize the bookmarks. That means, collected bookmarks can be annotated with concepts from the vocabulary (see Figure 7).

If users find a web resource, e.g. the manufacturer's website for a specific plastic or an article about a new material, they can pick it up into the shared bookmark collection and annotate it with concepts from the vocabulary, e.g. the specific plastic. If a needed concept does not exist in the shared vocabulary (e.g. in the case of a new material) or is not suitable (e.g. when the brand name changed), the users can modify an existing concept or add arbitrary tags. These new tags are automatically collected as "prototypical concepts" within the vocabulary. The users can consolidate them later by abstracting into concepts and placing them within the ontology. In this way, we allow that new concept ideas and tags are gathered seamlessly when they are occurring and that users can define concepts in a free and informal manner without the usual modeling overhead.

Figure 7: Tagging a web resource with concepts.

For structuring the concepts we concentrate on taxonomic relations because they are easy understandable for non-modelling experts. Going beyond common tagging systems, where the users can only bundle concepts underneath another, the users can organize their concepts within our approach in a shared structure according to the SKOS Core Vocabulary [3] in which the concepts are connected by "broader", "narrower" and "related" links. The users can further specify one preferred label and a number of alternative labels or synonyms (e.g. former brand names) as well as a textual description for each concept.

The users within the community share and maintain one taxonomy and one collection of bookmarks collaboratively. Everyone has the right for editing and modifications following the wiki paradigm of self-regulation.

In this way, we can support in particular the second and third transition phase of the ontology maturing process.

5.3 Tool support with SOBOLEO

We are developing SOBOLEO (**So**cial **Bo**okmarking and **L**ightweight **E**ngineering of **O**ntologies) in order to satisfy these requirements. SOBOLEO's goal is to support knowledge workers working together in one domain in developing a shared vocabulary and a shared collection of relevant web resources with a lightweight ontology editor and an ontology-enabled social bookmarking system.

[4] `http://www.im-wissensnetz.de/`

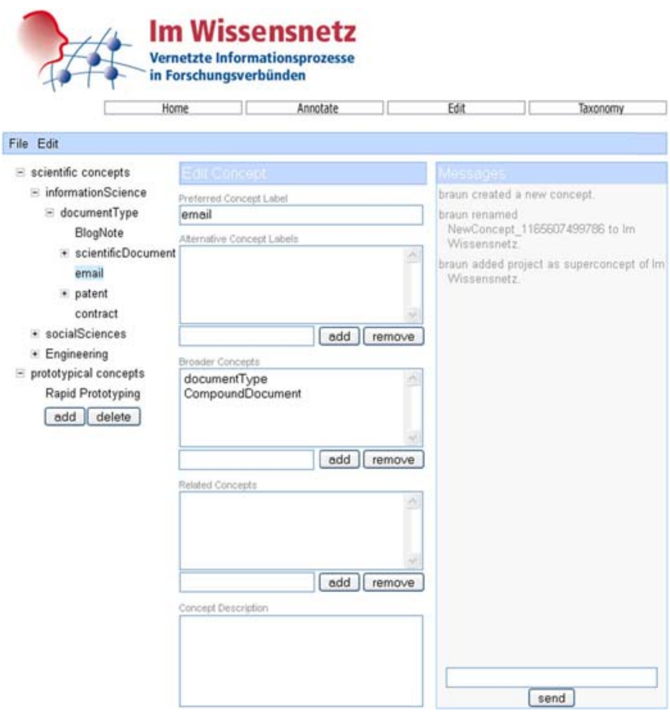

Figure 8: Ontology editor user interface.

Figure 9: Searching the bookmark collection.

SOBOLEO is based on AJAX technology using the Google Web Toolkit [11] and works in most current browsers—thus does not require any local installation. It consists of four application parts: an editor for the modification of the shared taxonomy, a tool for the annotation of web resources, a semantic search engine for the annotated web resources and a taxonomy browser for navigating the taxonomy and the bookmark collection.

- **Editing.** The editor interface (see Figure 8), built up tripartite, displays a tree view of the taxonomy on the left hand side of the screen. It shows the concepts with their preferred labels and their narrower and broader relations. Informal, not yet consolidated tags are collected in the special branch of "prototypical concepts". When a concept is selected in the tree view, its details are displayed in the center part of the screen. Here, the users can edit the preferred and alternative labels, the description, and the narrower, broader and related relations between concepts, which is further supported through auto completion of entities in the taxonomy. The right hand side of the screen provides a chat panel and allows having a conversation with other people editing the same ontology at the same time. The chat panel is also used for displaying changes made to the taxonomy. The system automatically generates a chat message that details who did which change. Changes done to the ontology by one user are visible almost instantaneously on all machines without requiring any intervention by the users. Updates of other users are also immediately reflected in the center screen part showing the details of the currently selected resource—particularly important if two users are editing one concept at the same time.

- **Annotation.** The annotation interface opens a pop up (see Figure 7) in which users can enter title and url of the web resource they want to add to the collection. Users can further specify concepts from the taxonomy, which is also supported by auto completion, or they can add new tags to annotate the resource. If the collection already contains the resource, the existing annotations are displayed and can also be edited.

- **Search.** The semantic based search engine allows for searching and retrieving resources within the shared bookmark collection (see Figure 9). Users can enter concept labels from the taxonomy or arbitrary search terms. The engine presents all resources either annotated with the identified concepts and their narrower ones or containing the input terms within the page content. The interface lists the resources with their title linking to the original page, with annotated concepts, a short excerpt of the page content highlighting the search terms, and the exact url. It further provides query refinement and relaxation proposals. Via each result's edit link, users can modify or remove the annotations of a web resource.

- **Browsing.** With the browsing interface users can navigate through the taxonomy and associated bookmark documents (see Figure 10). Starting from the root concepts, the users can click through the taxonomy concepts. On top, the users see the currently selected concept with its preferred and alternative labels and its description. Additionally, all its broader, narrower and related concepts are displayed as links for further navigation. Underneath the concept details there is a list of all resources which are annotated with the currently selected concept or with one of its

Figure 10: Browsing the taxonomy and associated bookmarks.

narrower concepts. These resources are further ranked by their date they were collected, thus the newest resources appear upmost.

6. CONCLUSIONS

We will only ever achieve sustainable ontology-based systems by embedding the task of building and maintaining ontologies into everyday work processes, enabling domain experts to do it without the help of knowledge engineers and by making it truly collaborative. We also have to acknowledge that ontologies cannot be formalized from scratch, but rather continuously evolve in a maturing process from informal tags to formal taxonomy hierarchies for which the ontology maturing process was presented.

Therefore, our model for ontology maturing offers four different steps. It allows for the emergence of ideas from each individual and the consolidation in communities for a common terminology. Then, in the third step of our model, relations help in creating formal lightweight ontologies. Finally, the fourth step could allow axiomatization. Such a maturing view on ontology engineering can overcome the problem of conceptual dynamics (e.g. the problem of the time lag between emergence of topics and their inclusion in an ontology).

In order to support such a maturing process, we presented two lightweight, easy-to-use and work embedded tools that allow the collaborative maturing of ontologies. Both of them lower the barriers to ontology editing for non-knowledge formulation experts. The project *IMAGINATION* will use the idea of image based ontolgy maturing with imagenotions for the collaborative creation of required domain ontologies. SOBOLEO allows social bookmarking with ontologies in the project *Im Wissensnetz*.

Our next work steps are evaluations and refinements of our model for the ontology maturing processes in these projects. We are mainly interested in evaluating which further functionalities are needed and wanted by the users, e.g. whether and when it is necessary to introduce big numbers of relations between the available concepts. Likewise, for the sake of simplicity, we are not considering aspects of ontology versioning at the moment (here, we want to refer to the works of [26]). Our main evaluation target is to analyze if the shared ontologies, developed based on our tools and models, converge to a stable and common accepted understanding for the available concepts over time. If not, our models require further support for agreement finding (e.g. like wiki) or personal ontologies.

7. ACKNOWLEDGMENTS

This work was co-funded by the European Commission within the project IMAGINATION, by the German Federal Ministry for Education and Research within the project Im Wissensnetz and by German Federal Ministry of Economic and Technology within the project KSIunderground.

8. ADDITIONAL AUTHORS

Additional authors: Gabor Nagypal (disy Informations-Systeme, Karlsruhe, Germany, email: `nagypal@disy.net`) and Valentin Zacharias (FZI Research Center for Information Technologies, email: `zach@fzi.de`).

9. REFERENCES

[1] H. Allert, H. Markannen, and C. Richter. Rethinking the Use of Ontologies in Learning. In M. Memmel and D. Burgos, editors, *Proceedings of the 2nd International Workshop on Learner-Oriented Knowledge Management and KM-Oriented Learning (LOKMOL 06), in conjunction with the First European Conference on Technology-Enhanced Learning (ECTEL 06)*, pages 115–125, October 2006.

[2] K. Barker, V. K. Chaudhri, S. Y. Chaw, P. Clark, J. Fan, D. Israel, S. Mishra, B. W. Porter, P. Romero, D. Tecuci, and P. Z. Yeh. A question-answering system for AP Chemistry: Assessing KR&R technologies. In *Proceedings of the Ninth International Conference on Principles of Knowledge Representation and Reasoning*, pages 488–497, 2004.

[3] D. Brickley and A. Miles. SKOS Core Vocabulary Specification. W3C working draft, W3C, November 2005.

[4] I. P. T. Council. The IPTC Standard, 2007.

[5] Del.icio.us. Del.icio.us. http://del.icio.us/, 2007. (accessed 2007-01-25).

[6] M. Denny. Ontology Tools Survey, Revisited. Technical report, XML.com, July 2004.

[7] M. Fernández-López and A. Gómez-Pérez. A survey on methodologies for developing, mainaining, integrating, evaluating and reengineering ontologies. Deliverable 1.4, EU IST Project IST-2000-29243 OntoWeb, 2002.

[8] Flickr. Welcome to Flickr - Photo Sharing. http://www.flickr.com/, 2007. (accessed 2007-01-25).

[9] S. Golder and B. A. Huberman. The structure of collaborative tagging systems. *Journal of Information Sciences*, 32(2):198–208, 2006.

[10] A. Gómez-Pérez, M. Fernández-López, and O. Corcho. *Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Advanced Information and Knowledge Processing. Springer, 1st edition, 2004.

[11] Google Web Toolkit. Google Web Toolkit - Build AJAX apps in the Java language. http://code.google.com/webtoolkit/, 2007. (accessed 2007-01-25).

[12] S. W. I. Group. Semantic Wiki State Of The Art, 2007.

[13] M. Guy and E. Tonkin. Folksonomies: Tidying up tags? *D-Lib Magazine*, 12(1), January 2006.

[14] M. Hepp. Possible ontologies: How reality constraints building relevant ontologies. *IEEE Internet Computing*, 11(1):90–96, January/February 2007.

[15] M. Hepp, D. Bachlechner, and K. Siorpaes. OntoWiki: community-driven ontology engineering and ontology usage based on Wikis. In *WikiSym '06: Proceedings of the 2006 international symposium on Wikis*, pages 143–144, New York, NY, USA, 2006. ACM Press.

[16] A. Hotho, R. Jschke, C. Schmitz, and G. Stumme. BibSonomy: A Social Bookmark and Publication Sharing System. In A. de Moor, S. Polovina, and H. Delugach, editors, *Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures*, Aalborg, Denmark, July 2006. Aalborg University Press.

[17] K. Kotis, G. A. Vouros, and J. P. Alonso. HCOME: A Tool-Supported Methodology for Engineering Living Ontologies. In C. Bussler, V. Tannen, and I. Fundulaki, editors, *Semantic Web and Databases. Second International Workshop - SWDB 2004*, volume 3372 of *LNCS*, pages 155–166, Berlin Heidelberg, Germany, August 2004. Springer-Verlag.

[18] C. C. Kuhlthau. *Seeking Meaning: A Process Approach to Library and Information Services.* Libraries Unlimited, Westport, CT, 2nd edition edition, 2004.

[19] A. Maedche, B. Motik, and L. Stojanovic. Managing multiple and distributed ontologies on the Semantic Web. *The VLDB Journal*, 12(4):286–302, November 2003.

[20] C. Marlow, M. Naaman, D. Boyd, and M. Davis. Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead. In *Collaborative Web Tagging Workshop at WWW2006*, May 2006.

[21] A. G. Perez, J. Angele, M. F. Lopez, V. Christophides, A. Stutt, and Y. Sure. A survey on ontology tools. Deliverable 1.3, EU IST Project IST-2000-29243 OntoWeb, 2002.

[22] Protégé. The Protégé Ontology Editor and Knowledge Acquisition System. http://protege.stanford.edu/, 2007. (accessed 2007-01-25).

[23] S. Schaffert. IkeWiki: A Semantic Wiki for Collaborative Knowledge Management. In *1st International Workshop on Semantic Technologies in Collaborative Applications (STICA'06)*, Manchester, UK, June 2006.

[24] A. Schmidt. Knowledge Maturing and the Continuity of Context as a Unifying Concept for Knowledge Management and E-Learning. In *Proceedings of I-KNOW '05, Special Track on Integrating Working and Learning*, July/August 2005.

[25] S. Sen, S. K. Lam, A. M. Rashid, D. Cosley, D. Frankowski, J. Osterhouse, F. M. Harper, and J. Riedl. tagging, communities, vocabulary, evolution. In *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 181–190, New York, NY, USA, 2006. ACM Press.

[26] L. Stojanovic. *Methods and Tools for Ontology Evolution*. PhD thesis, University of Karlsruhe (TH), Germany, 2004.

[27] J. W. Tanaka and M. Taylor. Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitve Psychology*, 23(3):457–482, July 1991.

[28] M. Völkel, M. Krötzsch, D. Vrandecic, H. Haller, and R. Studer. Semantic Wikipedia. In *Proceedings of the 15th international conference on World Wide Web (WWW'06)*, pages 585–594. ACM Press, 2006.