KNOWTA: Wiki-Enabled Social Tagging for Collaborative Knowledge and Experience Management

Martin Atzmueller[1], Fabian Haupt[1], Stephanie Beer[2], and Frank Puppe[1]

[1] University of Würzburg,
Department of Computer Science,
Am Hubland, 97074 Würzburg, Germany
{atzmueller,fhaupt,puppe}@informatik.uni-wuerzburg.de
[2] University-Hospital Würzburg,
Department of Gastroenterology,
Josef-Schneider-Str. 2, 97070 Würzburg, Germany
beer_s@klinik.uni-wuerzburg.de

Abstract. Social tagging systems are a useful tool for collaborative knowledge management: They enable the flexible collection, annotation, organization, and distribution of resources and information. A combination with wiki systems then provides powerful but easy to use approaches for a broad range of applications. This paper describes a wiki-enabled social tagging system for knowledge and experience management, and presents its application using a case-study in the medical domain.

1 Introduction

Both Web 2.0 and semantic technologies enable the creation of powerful tools for knowledge management. While social tagging systems [1] provide versatile approaches for the distributed collection, annotation, organization, and distribution of resources that are especially easy to use, wiki-based systems allow the collaboration of a community of users in a simple way.

Therefore, combining both approaches seems to be an attractive solution since they complement each other well: The wiki-based system provides the tools for accessing and managing the content, while the tagging application takes care of the annotation, and categorization of the resources. In this way, a community-driven knowledge management approach can be well supported. Additionally, by embedding semantic web [2] technology, e.g., for accessing the tagged resources or modeling the tag relations, a flexible and powerful user experience can be implemented.

This paper presents an approach combining tagging and wiki-editing in a social system for knowledge and experience management. The KNOWTA (*Know*ledge *Ta*gging) system provides advanced tagging functionality including recommendation options and utilizes a semantic wiki component for transparent access to the provided knowledge. KNOWTA is well suited for handling multi-modal knowledge, e.g., containing text and images. Both can be captured by the wiki engine and are transparently tagged (or semantically annotated) using the tagging and wiki functionality, respectively. We describe the system and its capabilities in detail and provide a real-world case study in the medical domain.

The rest of the paper is structured as follows: Section 2 outlines the basic issues of social tagging and the utilized semantic technologies. After that, Section 3 provides an overview on the proposed approach and discusses its elements in detail. Next, Section 4 presents a case study in the medical domain. Finally, Section 5 concludes with a summary and outlines interesting options for future work.

2 Social Tagging Systems

For organizing information and knowledge, the notion of *tagging* has recently received much attention: It has become a useful way for the collection, annotation, organization, and distribution of resources by users. The tags assigned to specific resources are used for navigation, locating the resources and for serendipitous browsing. Due to its ease of use, tagging systems provide an immediate benefit for users: They allow a transparent access to various resources. Additionally, tagging can help for communicating interesting nuggets of information [3].

In this paper we focus on the handling and management of multi-modal resources. We support both textual and image resources that are transparently integrated using (semantic) wiki technology. Therefore, we extend a conventional wiki system by providing extended means for the direct integration of images similar to the *Flickr* [4] system. Users can directly upload images as well as link other resources to the system. In the following, we shortly introduce the main features of social tagging systems, before we describe the KNOWTA system in the next section.

Social tagging systems are based on resources, e.g., bookmarks or images, users and tags. Thus, an entry e in such a system can be regarded as a 3-tuple composed of a resource $r \in R$, a user $u \in U$ and a set of tags $t_1, \ldots t_n \in T$, where R specifies the set of resources, U specifies the set of users and T specifies the set of valid tags.

In general, the set of tags T is unbound and can be extended by the users as needed. This is the case for folksonomies [5], for which arbitrary tags can be assigned. However, there are also other possibilities, especially for *closed communities*: In this case, the set of users U is fixed, and new users cannot join the community on their own. In such circumstances, often a restricted vocabulary is more appropriate. Then, the set T is prespecified by a *super-user* and cannot be freely extended. Usually, only a selected group of users, e.g., domain specialists, are enabled to modify this set of tags. In a knowledge-acquisition step, it is usually generated according to the specific domain and the targeted closed-community of users.

3 Wiki-enabled Social Tagging

This paper proposes the social tagging system KNOWTA for collaborative knowledge management. As a successor to the KNIZR(*Kn*owledge Orga*niz*e*r*) system [6], KNOWTA includes sophisticated semantic components for collaborative knowledge management. This section first provides an overview on the KNOWTA system. After that, the tagging and semantic capabilities are discussed in detail.

3.1 Overview on KNOWTA

KNOWTA builts upon established Wiki technology based on the *JSPWiki* system [3]; it supports a variety of resource types that can be embedded into a wiki page, in addition to the rich textual edit actions. Specifically, the system enables the management of image data with corresponding texts for knowledge management. Then, image and textual data, i.e., multi-modal information, can both be tagged and semantically annotated using the functionality of the social tagging system and/or the wiki-features, respectively. As an optional extension, the system can be configured for closed-communities such that controlled vocabularies for the tags, i.e., the tag ontology of the system can be defined. This is especially useful for specific application projects with a fixed specific set of contributing users.

In addition to the JSPWiki component, KNOWTA also utilizes the *semantic core* component of the *KnowWE* [7] system. The latter allows for powerful semantic browsing options for accessing the stored resources and content. For semantic browsing and querying users, resources, tags, links, and the content itself can be considered. In addition to including complex handcrafted ontologies, KNOWTA can also include SKOS [8], i.e., the *Simple Knowledge Organization System*, a W3C standard for knowledge organization for the web and especially the semantic web. SKOS is based on (simple) web-standards such as RDF, and provides a formal language for constructing, for example, thesauri or concept taxonomies. At its core it is quite simple, extensible and maintainable. Figure 1 shows an exemplary screenshot of the KNOWTA system.

For general resource management, KNOWTA offers the following functionality:

- Resources can be created by providing a title, an image, a set of tags and/or a description (optional), and a set of tags can be assigned to them based on a describing sentence. This is implemented using the wiki engine and the functionality of the tagging engine. Additionally, as described below, a set of tags can be semi-automatically assigned using recommended selections.
- Resources are defined for certain groups of users: Using an access control schema of JSPWiki, certain user groups and corresponding rights can be defined.
- Resources can be commented on by the users: This enables collaborative knowledge and experience management and can support the collaborative discussion of specific issues regarding the given resource.
- Resources and their associated (multi-modal) information can be semantically annotated: This is accomplished using the features of the (extended) semantic core of the KnowWE system. By adding semantic annotations to the textual information, links and relations between resources can be easily established.

Connections between resources/tags are thus either established using (explicit) links between the resources, implicit relations/links given by the a matching set of tags of different resources, or by considering the relations between the tags defined in a domain ontology (for closed communities). The general knowledge and experience management functions are supported by the wiki-enabled information handling – using unstructured, semi-structured, and structured information. Since the knowledge can be collaboratively extended, the system provides a knowledge-rich collaborative environment for a broad range of applications.

[3] `http://www.jspwiki.org`

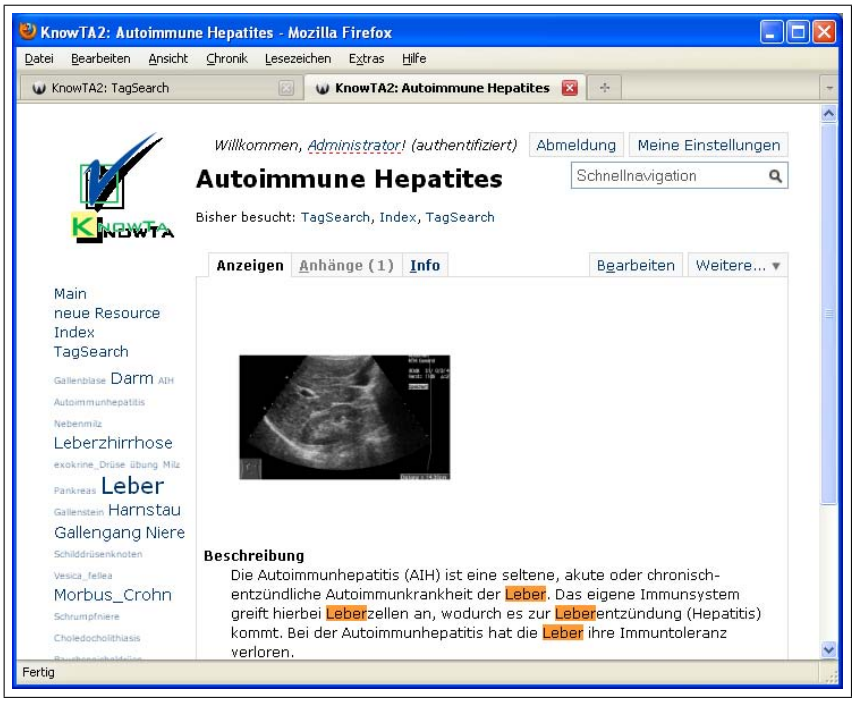

Fig. 1. KNOWTA: Exemplary resource/image with annotation and tags (in german).

3.2 KNOWTA– Tagging, Recommendation, Browsing

For the tag assignments, KNOWTA utilizes a tag recommendation system that takes a short sentence as input, and then proposes a set of suitable tags. This feature is especially important for the targeted medical domain, since the users often do not have a lot of time when entering a new resource, a situation which is quite typical, for example, for medical doctors. Figure 2 shows the simple interface for creating a new resource. Figure 3 shows an example of the medical domain (liver) demonstrating the inline tag edit functionality.

For the tag recommendation, text mining techniques are applied. There are several options for tag recommenders (see [5] for a survey); for occurrence-based tag recommendation the following strategy is applied: The tag assignment information of all users is taken into account for computing the co-occurrence counts of all tags, weighted by their importance, similar to a *tf/idf* schema [9]. After that, the input text (given by a short sentence, and/or a set of tags) is preprocessed by tokenization, removing stopwords, and stemming. Then, a set of tokens with the highest co-occurrence counts (and weights) is retrieved and proposed to the user. The user can then apply these tags and use them for tagging the current article. This schema allows providing a controlled vocabulary, since the 'free tagging' functionality can be disabled in favor of the tag rec-

Fig. 2. KNOWTA: Interface for creating a new resource (screenshot in german): The user can upload a new resource/image, provide a default tag/title (Seitenname), enter a description (Beschreibung) which is used for recommending tags, and enter a list of tags (Tags).

ommendation schema using default occurrence values. Then, only the given tags in the vocabulary can be proposed and used for tagging.

After a set of resources has been defined, a key point are efficient and effective search and browsing capabilities for these resources. According to Shneiderman [10] an effective user interface should implement features for getting an overview first, then zoom on some details, finally getting the details on demand. In this way, the so-called *visual information seeking mantra* is implemented. KNOWTA implements these features by both a powerful search facility and also by using a tag cloud visualization. Additionally, for each resource a set of related tags/resources is shown such that the navigation between them is easily implemented. Then, the user can apply a dynamic query refinement for obtaining a refined set of resources within a more detailed cluster of information.

3.3 KNOWTA– Semantic Representation

As mentioned before, some of the features of KNOWTA are embedded into the semantic core functionality of KnowWE which was extended as needed. The semantic core of KnowWE relies on Sesame[4] as RDF-Triple storage and OWLIM[5] as reasoning engine.

KnowWE (without considering any extensions) is a basic semantic wiki system featuring an annotation syntax and the possibility of SPARQL-query embedding ([7]). It provides several ways of expanding the ontology of the wiki: Ontologies can be introduced to the system, for example, by embedding owl-source directly into a wikipage within `<owlextension>...</owlextension` tags. For more extensive extensions of the ontology it is advisable to use the provided ontology manager interface to

[4] `http://www.openrdf.org/`
[5] `http://www.ontotext.com/owlim/`

Fig. 3. KNOWTA: Exemplary resource/image with tag-editing mode (in german).

upload owl-files directly into the wikisystem. Another common operation is the introduction of new annotation properties into the system, which is facilitated by the shorthand `<properties>...</properties>`. Properties can be listed within those tags anywhere in the wiki and are then recognized as viable annotation properties for any resources.

As the tagging of a resource is a specialized case of a semantic annotation, the tagging functionality is merely a convenient way to introduce this annotation into the backend ontology. Adding a tag to a resource can be done either by using a AJAX-based tag edit panel or by editing the content of the XML-Tag `<tags>...</tags>` directly. The former saves the tags into the XML tags and the wikisource as well. After updating a wikipage, KnowWE parses the page into a specialized datastructure, which in turn provides an interface to create owl statements from the wikiarticle. Tags are rendered into multiple N-ary statements as introduced in [7].

The following owl snippet, for example, is the result of tagging the page `Main` with the tag `Demo`, i.e., the tag is given in the wiki source by `<tags>Demo</tags>`, see Figure 4 for an example.

```
<rdf:Description rdf:about="lns#Main..Node9_XML_content..MainhasTagDemo">
        <rdfs:isDefinedBy rdf:nodeID="node41"/>
        <rdfs:seeAlso rdf:nodeID="node41"/>
        <rdf:predicate rdf:resource="ns#hasTag"/>
        <rdf:object rdf:resource="lns#Demo"/>
        <rdf:subject rdf:resource="lns#Main"/>
</rdf:Description>
```

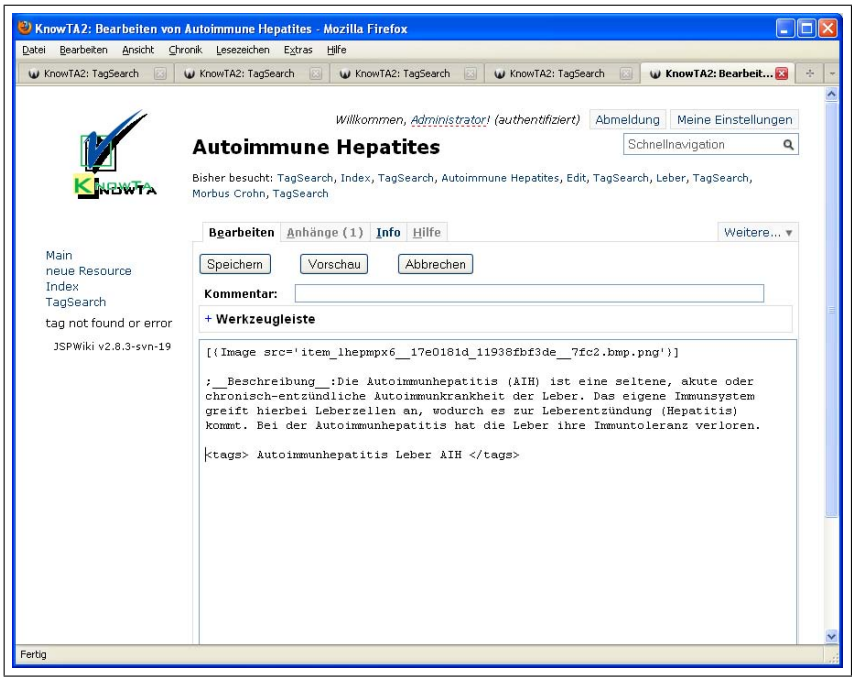

Fig. 4. KNOWTA: Exemplary article with associated tags (in german).

The namespaces used are `lns` which is the namespace of the local wiki installation and `ns` which stands for the KnowWE upper ontology namespace included in in KnowWE. `node41` is a reference to the node in the specialized datastructure in KnowWE. This allows, among others, a precise localization of the annotation, or in this case the tag element within the wiki. The name of the node is automatically generated by convention of the semantic engine.

Due to the fact, that the tags are stored in the ontology of the wiki, it is possible to integrate arbitrary SPARQL-queries for tags or pages tagged with specific tags everywhere in the wiki, for example, within any page or even the menu structure. To get a list of all pages, for example, with the assigned tag `Demo` the following SPARQL-query can be embedded into a wiki page. It results in a list of pages with the `Demo` type.

```
<sparql>
SELECT ?q
WHERE {
    ?t rdf:object lns:Demo .
    ?t rdf:predicate lns:hasTag .
    ?t rdfs:isDefinedBy ?o .
    ?o ns:hasTopic ?q
}
<sparql>
```

To get an overview of all tags used in a wiki, a tag cloud can be rendered anywhere in a wikipage (see figure 5). As expected, the size of the rendered tags corresponds to the

tag frequency within the whole wiki. Clicking on a tag searches for all pages that are tagged with the selected tag and allows a tag-based navigation through the wikispace.

Fig. 5. An exemplary tag cloud showing the tags of a wikispace.

4 Case Study

The case study presents an application of the KNOWTA system for extended knowledge and experience management. The application domain is the medical domain of sonography, for which sonographic images are collected, annotated and commented in order to serve as instructive examples for typical but also exceptional features of certain disorders. In this way, effective tutoring and discussion between the examiners can be initiated. So far, about 500 images have been obtained, which comprehensively cover the problem domain and provide an experience base for supporting medical training and consultation by knowledge and experience management. Additionally, the system provides the capability to semi-automatically export the collected data to the *CaseTrain* training system [11].

4.1 Knowledge and Experience Management

The context of the application is given by the intelligent documentation and consultation system SONOCONSULT system [12, 13] – a medical system for sonography. The system is in routine use in several hospitals, for example, in the DRK-hospital in Berlin/Köpenick and in the University Hospital of Würzburg.

For a sonography (ultrasound) examination, the examiners need to closely inspect the ultrasound images, for example, see Figure 6 in order to document the correct findings. Since this process is highly subjective and also significantly dependent on the experience of the examiner, there are often discrepancies between beginners and more senior examiners concerning the correct findings, c.f., [14]. Therefore, a system for the collection and annotation of *interesting*, i.e., exceptional or typical images for certain medical phenomena with associated textual descriptions provides for a powerful tool.

In this context, KNOWTA thus provides an ideal framework for implementing an image pool for tutoring, targeted training and general knowledge management. Special cases of ultrasound phenomena can be uploaded to the system. Figure 7 shows an example for a set of liver-related phenomena.

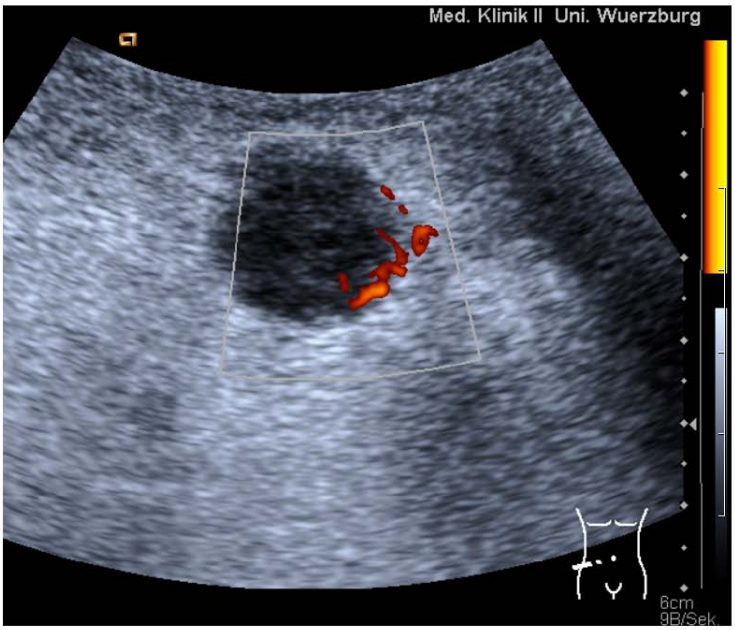

Fig. 6. Exemplary image from an ultrasound examination.

The resources can be easily described, tagged and annotated, and junior examiners can search for specific (difficult) situations using the tag cloud and/or the search functionality. Additionally, hints can be easily communicated using the annotation and/or comment function of the system. Furthermore, using the SKOS features of the system the terminology of the different organ systems and diseases can be directly utilized for an effective retrieval of resources.

4.2 Knowledge Capture for Tutoring

After the resources have been collected in the KNOWTA system, they are readily available for further processing. In order to apply the system for knowledge capture for designing and building tutoring cases, a (semi-)automatic export to the CaseTrain system is provided. Furthermore, an automatic option is given by exporting all images with their associated 'main' tag (the page title) and storing the complete list as 'long menu' questions for all images. In this case a (potentially randomly selected) subset of tagged images is obtained; for tutoring, the user needs to select a specific tag from the 'long menu' that describes the image best.

So, the images, the annotations and tags can be utilized for creating new quizzes for tutoring and teaching purposes. Additionally, the generated cases can be potentially linked to KNOWTA articles, such that CaseTrain users can get background information about the case. An exemplary CaseTrain screenshot is given in Figure 8.

Fig. 7. Exemplary tag-search for liver diagnoses (in german).

5 Conclusion

Social tagging systems can provide powerful and intuitive solutions to various knowledge management problems. In this paper, we have introduced the KNOWTA system for tutoring and general knowledge management. We discussed the different components and features of the system, and we have described its application in a real-world medical application scenarios.

So far, we have implemented a full prototype of the system: Since the results are quite promising, we aim to perform a comprehensive evaluation of the system in the sketched application scenario. Additionally, we want to utilize as much background knowledge as possible in order to exploit the semantic features at their full level. We also plan to enhance the tag recommendation feature by discovering certain user subgroups and applying their features for a better recommendation, knowledge capture, and user experience using data mining [15], text mining [16] and information extraction [17] techniques.

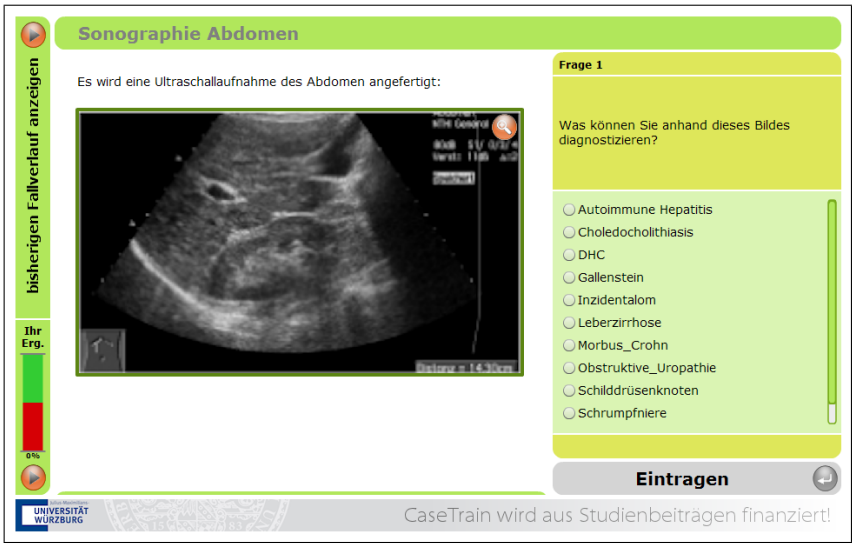

Fig. 8. Exemplary CaseTrain case.

Acknowledgements

This work has been partially supported by the German Research Council (DFG) under grant Pu 129/8-2.

References

1. Golder, S., Huberman, B.A.: Usage Patterns of Collaborative Tagging Systems. Journal of Information Science **32**(2) (April 2006) 198–208
2. Antoniou, G., Harmelen, F.v.: A Semantic Web Primer. MIT Press, Cambridge, Mass. (2004)
3. Hammond, T., Hannay, T., Lund, B., Scott, J.: Social Bookmarking Tools (I) - A General Review. D-Lib Magazine **11**(4) (April 2005)
4. Mislove, A., Koppula, H.S., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Growth of the flickr social network. In: Proceedings of the 1st ACM SIGCOMM Workshop on Social Networks (WOSN'08). (August 2008)
5. Jäschke, R., Marinho, L.B., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in folksonomies. In: Proc. PKDD '07. Volume 4702 of Lecture Notes in Computer Science., Berlin, Heidelberg, Springer (2007) 506–514
6. Atzmueller, M., Hörnlein, A.: Exploiting the Power of Social Tagging Systems: A Semantic Flickr Approach for Tutoring and Knowledge Management. In: Proc. 1st European Workshop on Design, Evaluation and Refinement of Intelligent Systems, Erfurt (2008)
7. Reutelshoefer, J., Haupt, F., Lemmerich, F., Baumeister, J.: An extensible semantic wiki architecturre. In: SemWiki'09: 4th Workshop on Semantic Wikis. (2009)
8. Brickley, D., Miles, A.: SKOS core vocabulary specification 2005-11-02. W3C working draft, W3C (November 2005) http://www.w3.org/TR/swbp-skos-core-spec.

9. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. 1st edn. Addison Wesley, Harlow (1999)
10. Shneiderman, B.: Direct Manipulation for Comprehensible, Predictable and Controllable User Interfaces. In: Proc. 2n Intl. Conf. on Intelligent User Interfaces. (1997) 33–39
11. Hoernlein, A., Ifland, M., Kluegl, P., Puppe, F.: CaseTrain: Design and Evaluation of a Case-Based Training-System and its Comprehensive Application in University (Konzeption und Evaluation eines fallbasierten Trainingssystems im universitätsweiten Einsatz (CaseTrain)). GMS Med Inform Biom Epidemiol **5**(1) (2009)
12. Huettig, M., Buscher, G., Menzel, T., Scheppach, W., Puppe, F., Buscher, H.P.: A Diagnostic Expert System for Structured Reports, Quality Assessment, and Training of Residents in Sonography. Medizinische Klinik **99**(3) (2004) 117–122
13. Puppe, F., Atzmueller, M., Buscher, G., Huettig, M., Lührs, H., Buscher, H.P.: Application and Evaluation of a Medical Knowledge-System in Sonography (SonoConsult). In: Proc. 18th European Conference on Artificial Intelligence (ECAI 20008). (2008) 683–687
14. Atzmueller, M., Puppe, F., Buscher, H.P.: Profiling Examiners using Intelligent Subgroup Mining. In: Proc. 10th Intl. Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP-2005), Aberdeen, Scotland (2005) 46–51
15. Atzmueller, M., Puppe, F.: A Knowledge-Intensive Approach for Semi-Automatic Causal Subgroup Discovery. In: Knowledge Discovery Enhanced with Semantic and Social Information. Springer Verlag (2009)
16. Atzmueller, M., Nalepa, G.J.: A Textual Subgroup Mining Approach for Rapid ARD+ Model Capture. In: 22nd International Florida Artificial Intelligence Research Society Conference (FLAIRS). (2009) 414–419
17. Atzmueller, M., Kluegl, P., Puppe, F.: Rule-Based Information Extraction for Structured Data Acquisition using TextMarker. In: Proc. LWA 2008 (Knowledge Discovery and Machine Learning Track), University of Wuerzburg (2008)