Enabling Knowledge Workers to Collaboratively Add Structure to Enterprise Wikis

Florian Matthes, Christian Neubert
Technische Universität München, Munich, Germany
matthes@in.tum.de
neubert@in.tum.de

Abstract: Varied fields of application, fast access to often needed information, easy collaboration capabilities and low maintenance costs make wikis very attractive for enterprises. For these reasons in many companies wikis have already been firmly established as tools for collaboration and knowledge exchange.

Since most of the content in wikis is completely unstructured (plain hypertext, links, etc.) it is difficult for programs to process the information on the particular wiki pages. Therefore individual pages can only be found by means of a full-text search engine, but searching for particular pages with specific attributes and attribute values is not possible.

In this paper we present Hybrid Wikis, a lightweight approach for structuring content and management of information structures in enterprise wikis. Hybrid Wikis are realized based on the commercial Enterprise 2.0 software Tricia and supported by our experiences made with classical wikis, semantic wikis and integrated Enterprise 2.0 platforms used for knowledge and information management in enterprises.

Inspired by these web technologies Hybrid Wikis extend the wiki provided by Tricia with a few mechanisms for classification, linking, consistency checking, and visualization of wiki pages, which can be combined flexibly.

We explain how these mechanisms facilitate the structuring of content in enterprise wikis and how both can benefit from it, knowledge workers and enterprises. Hybrid Wikis create incentives for users to apply structure by giving suggestions of frequently used structured elements, provide lightweight web-interfaces which enable users to manage the structured elements directly as part of the page content, and help to avoid information redundancies by offering structured searches as well as autocompletion mechanisms for structured elements. Furthermore we show how Hybrid Wikis enable knowledge workers to manage and integrate structured and unstructured information uniform across the enterprise, which is one of the key challenges knowledge management systems are faced with.

Keywords: Enterprise 2.0, Social Software, Semantic Web, Wikis, Structuring of Content, Knowledge Management Systems

1. Motivation and Overview

1.1 Learning from the Wikipedia Encyclopedia

The idea of the Wikipedia (Leuf 2001) is quite simple: Collecting and providing knowledge as an encyclopedia based on an open Internet platform freely available for all users in the web. Each particular (wiki) page describes an article about a narrowly defined topic. A main characteristic of the Wikipedia is that users directly can edit the pages and immediately contribute to the content. In contrast to classical content management systems, in wikis changes made by the users are immediately visible for other users without having the changes be revised. Thus, wikis make it possible for information consumers, e.g., readers of a topic, to contribute their knowledge very quickly (wikiwiki = quick, Hawaiian). Therefore the boundary between information consumer (reader) and producer (editor) becomes increasingly blurred. In (Chang 2006) this phenomena is called *Prosumer*. Furthermore wikis are easy to use, since the editors normally do not need system instructions for applying their contributions. Even if only a small number of readers contribute to the community, i.e., write contributions or make corrections, a huge amount of knowledge and topics has been collected in the last years.

Since different editors contribute to the same topic the content in wikis is conjointly created and revised by all participants (Surowiecki 2004). With the increasing number of topics and complexity in the Wikipedia project structures (categories, folder, formatting conventions, etc. …), processes (discussion, consolidation, etc. …) and roles (user groups with different access rights) emerged over time in order to ensure the quality of the encyclopedia. In the Wikipedia authors are motivated to share their competences and concepts with other users in order to get feedback by means of page view counts and backlinks to their contributions. Besides the authors, moderators and users reorganizing the contents play another very important role in the project, since they are responsible

for the linguistical and structural consistency between different topics as well as for the categorization and linking of the content. Those users are rather different to the page authors.

The description of each topic is primary effected in a textual form and can be complemented by images or embedded media objects (e.g., videos). Furthermore all topics can be set in relation to each other by means of hyperlinks. Thus over time a very complex net of information emerges, which is primary used by readers to navigate and search in this information space. Moreover links also enable search engines or visualization tools to automatically explore and process the wiki contents.
Besides the unstructured mainly textual described topics the pages in the Wikipedia project are complemented by semi-structured elements. For instance, typical attributes of a topic are shown as key-value-pairs in a tabular view on the wiki page, e.g., for the topic Passau: Population = 50.627.

The Wikipedia provides the concept of templates (Haake 2005) in order to enable authors of wiki pages to conserve often used structures and descriptions form the same class of elements (e.g., the structure of a page describing a town) and to reuse this structure in a similar context (e.g., describing another town). One the one hand templates define the layout of a wiki page and on the other hand the structure of the wiki page content. For example, a template representing a town topic could specify, that the template always includes two sections describing the history and the climate as well as an attribute representing the population, which is shown in a tabular view at the right margin. Wikipedia templates are created independent of the particular wiki pages, i.e., authors manually derive them from patterns observed on the different pages. Similarly to the wiki page contents the templates are conjointly developed by the template authors. Furthermore, templates are also freely available for all authors. One the one hand templates facilitate the reuse of existing structures, one the other hand they also represent a jointly produced consensus of the authors for a particular concept, e.g., the description of a town should always contains the attributes *Population* and *Headquarters*. Therefore, by using templates an implicit conceptualization takes place, i.e., a typing of the particular wiki pages.

1.2 Enterprise Wikis & Social Software
Motivated by the success in the Internet wikis are firmly established as tools for communication and collaboration tasks in enterprises today (Enterprise 2.0 Fallstudien - Aus Erfahrung lernen Webseite 2011). Since wikis are universally applicable in different enterprise contexts (e.g., the documentation of complex business processes), they are very attractive for enterprises. Besides wikis further means of communication frequently used in the Internet are applied in enterprise contexts. For instance, blogs and microblogs are used to interchange enterprise-relevant news or expert knowledge is exchanged across the departmental borders via social enterprise networks. The application of theses web techniques and technologies in companies is called Enterprise 2.0 (McAfee 2005). By means of social interaction as well as incremental, decentralized extension and cross-linking emergent structures for information, functions and processes arise, which are often superior to classical standard solutions for information and knowledge management in enterprises.

A great number of software vendors combine different individual Web 2.0 (O'Reilly 2008) solutions to integrated Enterprise 2.0 platforms which are especially optimized for their application in the enterprise (T. Büchner 2009). Besides the advantages of Web 2.0 applications (easy to use, support of multimedia objects, management of revisions, discussions, scalability, open programming interfaces) these platforms provide extended search capabilities, tagging, access control and web-interfaces, which are similar to desktop applications. Since users and groups in enterprises are more coherent and the number of collaborating users in companies is much smaller than in the Internet, further additional (wiki-) services are supported, e.g., alerting (e-mail notification in case of changes in the wiki page content), awareness (individual filters concerning all changes in the platform). A prominent example for an integrated Enterprise 2.0 platform is the Microsoft Share Point 2010 (Microsoft SharePoint 2010 Webseite 2010).

2. Problem Statement - Semantic Wikis: From Unstructured to Structured Content
Since contents in (enterprise) wikis mainly consist of unstructured elements (hypertext, links), for programs is very difficult to automatically process the information on the wiki pages. Thus, it is only possible to find individual topics by means of their textual descriptions, but a selective search for a topic with a specific attribute value is not possible, e.g., it is not possible to define a search for towns with a population greater than 500.000.
Semantic web technologies are targeting this problem by providing authors mechanisms to enrich the page contents with additional information (metadata). These metadata are not visible for human users

and can be processed and understand and by programs (Semantic MediaWiki Webseite 2006). Wikipedia templates are rather targeting the author needs (e.g., reuse of structures) and types as well as information models emerge implicitly by applying them to particular wiki pages. In contrast to templates in semantic wikis the typing of wiki pages as well as the definition of attributes and relations explicitly takes place. The definition of types, attributes and relations in semantic wikis is often based on a formal language which is processable by other programs. In order to annotate wiki pages with metadata the authors are either supported by a tool or they have to append the metadata manually by means of a specific wiki language (Wiki markup 2011).

Thus, in semantic wikis the pages and its metadata correspond to an explicitly formulized information model of the authorship and the schema designers[1]. In order to enable programs to define and access these metadata uniformly via the Internet (e.g., semantic search engines) a great number of semantic wikis implement the standardized Resource Description Framework (Resource Description Framework (RDF) 2011). By using a standardized language for the definition of resources by means of RDF it is possible to define very specific queries to find relevant elements in the universe of the available RDF sources in the Internet.

Semantic technologies and semantic wikis already exist for many years. However, they are still less common in the Internet as well as in enterprises. We assume that this is for the following reasons:
- Authors have additional expenditure when annotating the contents with metadata
- The tools and wiki languages provided for the management of the annotations are unsuitable for the authors
- Authors do not immediately benefit from the annotations (metadata)
- The modeling elements provided by the RDF are difficult to understand for authors
- The schema the metadata are based on is often created by an (ontology) engineer in a separate process independent of the wiki page authors and contents. Therefore the underlying domain is not understandable and transparent for the authorship

In the following we introduce a new approach, called Hybrid Wikis, facilitating the structuring of content in an enterprise wiki. We introduce a few concepts and mechanisms and subsequently explain how they contribute to an improvement of the aforementioned reasons.

3. Hybrid Wikis - Enable Knowledge Workers to Collaboratively Add Structure to Content
The concept of Hybrid Wikis is developed since 2009 and based on the experiences made with classical wikis, semantic wikis and integrated Enterprise 2.0 platforms. Hybrid Wikis are realized based on the commercial Java-based Enterprise 2.0 software Tricia (Tricia - Hybrid Wikis for Collaboration and Information Management Webseite 2010) providing a generic framework for the development of Web 2.0 applications (Büchner 2010). Thereby the Tricia Enterprise 2.0 software is extended by a few mechanisms for classification, linking, consistency checking, and visualization of wiki pages, which can be combined flexibly.

Similar to traditional enterprise wikis in Hybrid Wikis authors can easily and quickly modify unstructured content, but at the same time they are also empowered to structure the content of wiki pages in the same manner (easy and quick). Hybrid Wiki provide a simple way for structuring contents without the authors have to learn a specific (wiki) syntax for annotating metadata or to use a specialized tool for the definition of a formal description language for resources (e.g., RDF). Therefore, Hybrid Wikis lower the technical hurdle for authors in order to provide structured data. Hybrid Wikis work as described in the following subsections.

3.1 Attributes and Values
Besides the unstructured content for each wiki page a table consisting of simple key-value-pairs (attribute-value-pairs) can be created (c.f., Figure 1). In the contrast to template-based wiki systems (c.f., Section 1.1 Learning from the Wikipedia Encyclopedia) in which the attribute definitions are statically bound to the corresponding template, in our approach the attributes (with corresponding values) can arbitrarily be assigned to the wiki pages, i.e., the attributes in Hybrid Wikis are independent of the underlying templates. An attribute value can either be a link or a plain string.

---

[1] The schema designers provide the underlying ontology

Figure 1: Wiki page with unstructured and structured (type tags, attributes, values) content.

3.2 Type Tags

Typing of wiki pages takes not place implicitly by the usage of templates, in Hybrid Wikis the type of a page is determined by so called type tags. A type tag is a keyword (Golder 2005) with a special semantic (the type of the page), which can be explicitly attached to the wiki page by a user. For each wiki page an arbitrary set of type tags can be assigned, i.e., it is possible that a wiki page belongs to more than one type. Besides the type tags, conventional tags (e.g., todo, interesting) can also be assigned to the pages. In Figure 1 the conventional tags *rss* and *feed* as well as the type tags *issue* and *feature request* are used. When an user clicks on a type tag, the table as illustrated in Figure 2 is shown. This table lists all instances of the corresponding type ordered by the frequency of non-empty attributes (i.e., non-empty attribute values). As shown in Figure 1 the (type) tags as well as attributes and values are directly available as part of the page content. Therefore in our approach typing is supported by editing the page content and not by manipulating the underlying markup.

Versions   New Edit Delete

Wiki Pages with Type Tag feature request in Tricia - Issues

Tags: no tags assigned   edit tags

Text...

Showing 1 to 10 of 177 entries   Search: First Previous 1 2 3 4 5 Next Last

	status (175)	priority (166)	assigned_to (161)	effort (108)	requested_by (92)	Installation (28)	Submission Date (19)	Feedback Release (19)	Url (19)
"New"-Button in der Suche (auch embedded searches)	open	medium	lex	high					
3060 MAXIMAL_ONE sollte "at most one value" heissen	open	low	Chris	low		localhost	2011.01.16 - 13:15:45	3.0.1_02fcba	http://131.159.30.56:8091 /typeTag /instances?id=root
Access rights string in admin view	open	low	Tricia Team	low					

Figure 2: Tabular overview of all wiki pages with type tag *feature request*.

3.3 Suggestions and Autocompletion

Based on a statistical analysis of frequently used combinations of tags, type tags and attributes in Hybrid Wikis type and attribute suggestions are generated. In Figure 1 the attribute *Installation* is suggested to be filled out by the author. Thereby one the one hand structuring of content is facilitated on the other hand the authors are urged to reuse frequently used and well-established structures. In our approach templates are not created in a separate process derived from the author's observations (pattern), but they rather emerge from the management and structuring of the contents of particular wiki pages, their tags and by means of an intelligent suggestion mechanism. Furthermore templates can be derived by means of an analysis of the data set (wiki pages) in the wiki system. By providing type and attribute suggestions the authors are urged but not forced to contribute structure. In our approach the information model (domain model) emerges by the management of contents, i.e., the structures and properties of the model arise spontaneously due to interplay of the particular wiki pages. The advantage of this approach is that the underlying information model (with its concepts, attributes, and relations) emerges from the collaboration in the wiki and therefore mainly represents the consensus of the entire authorship.

Furthermore the editing process of type tags, attributes and values is assisted by autocompletion results (c.f., Figure 3). Thereby the reuse of frequently used structure is facilitated and the terminology of the wiki community converges toward a certain set of terms and elements, the domain language.
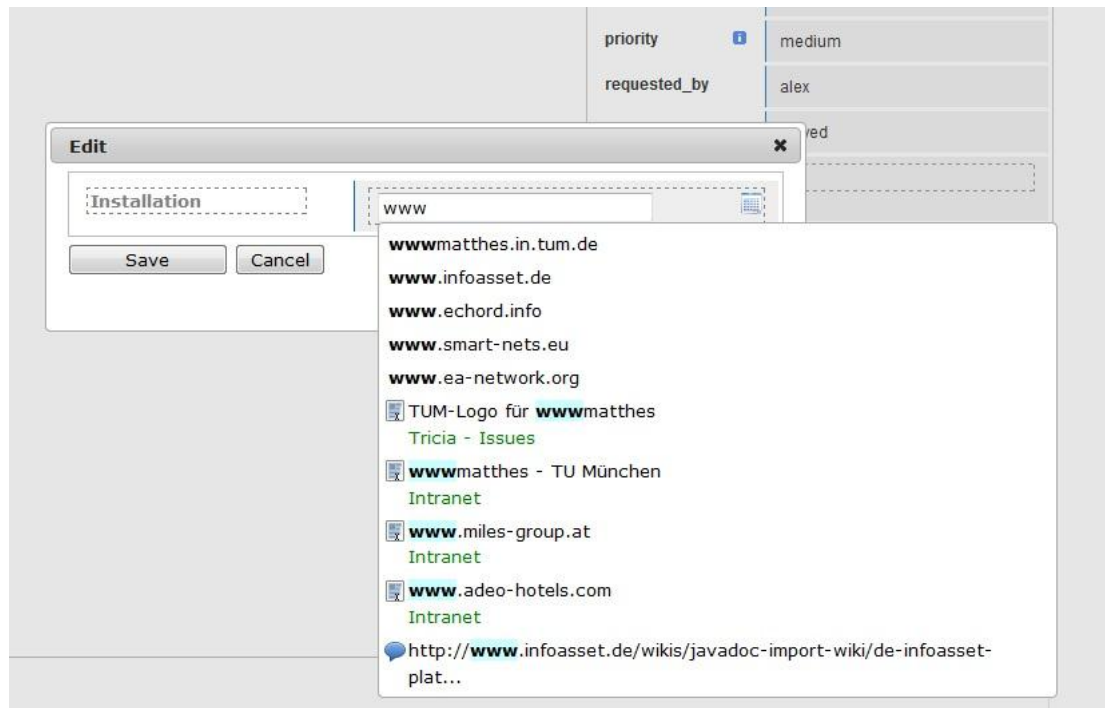
Figure 3: Autocompletion with highlighting for values of the attribute *Installation*.

3.4 Constraints

In addition to the possibility to directly and dynamically derive the information model from the wiki pages (bottom-up modeling) in Hybrid Wikis the state of a model can be defined in a more specific way. It is possible to bind attributes to a conceptual definition, which is called type tag definition. By doing so, integrity constraints can be specified for attributes, e.g., the type and the range of the values of an attribute. For instance, in Figure 3 the range of attribute values is specified as {open, rejected …}. Furthermore concepts and attributes explicitly can be specified, i.e., without the existence and analysis of a certain set of wiki pages.

Type tag definitions and attribute definitions enable users to create information models (concepts, attributes, and relations) in a more conventional way, top-down.
Type tag definitions are not separated from the data, they immediately have an impact on the particular wiki pages. For example, if an attribute definition *status* for type tag definition *bug* is specified, on all pages having the type tag *bug* a corresponding attribute suggestion is shown. Thus, the decisions made by a schema designer are immediately visible for the authorship.
Furthermore the constraints influence the autocompletion results for attribute values. For instance, if the range of an attribute is specified as {open, rejected …}, these values are preferred in algorithm which is calculating the autocompletion results. Since users are never forced to enter a specific value constraints in Hybrid Wikis are called soft constraints.

Beside currently assigned attribute definitions the dialog for the management of a type tag definition also shows attributes which are used in combination with the corresponding type tag (c.f., Figure 2). Additionally for all attributes it is shown how frequently they are used. In this way even the schema designer receives attribute suggestions and she could also identify candidates being adopted in schema (type tag definition). Furthermore she is able to reflect the quality of her decisions with regard to the schema by checking the usage of currently bound attributes.
By providing attribute suggestions for both, particular wiki pages and the schema management dialog, model designer (ontology engineers) and domain experts (wiki page authors) certainly enter into dialogue. Hybrid Wikis support building of consensus in both directions, from the data to an adapted type definition and from a given type definition to a revision of incompatible or missing values.
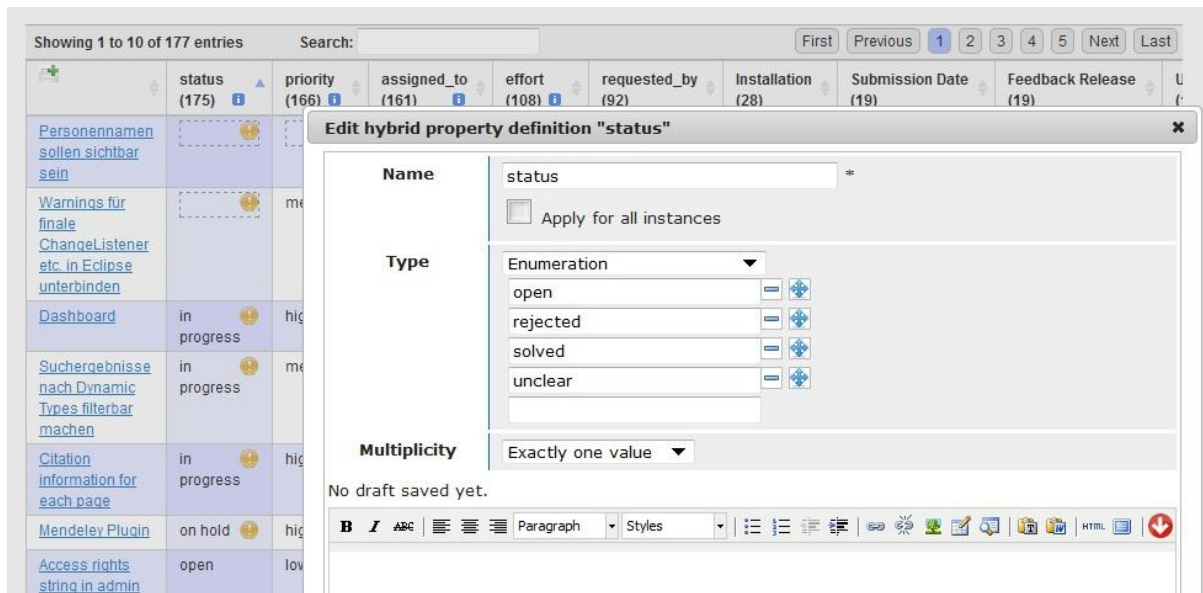
Showing 1 to 10 of 177 entries | Search: | First | Previous | 1 | 2 | 3 | 4 | 5 | Next | Last

	status (175)	priority (166)	assigned_to (161)	effort (108)	requested_by (92)	Installation (28)	Submission Date (19)	Feedback Release (19)	U
Personennamen sollen sichtbar sein									
Warnings für finale ChangeListener etc. in Eclipse unterbinden		me							
Dashboard	in progress	hig							
Suchergebnisse nach Dynamic Types filterbar machen	in progress	me							
Citation information for each page	in progress	hig							
Mendeley Plugin	on hold	hig							
Access rights string in admin	open	lov							

Edit hybrid property definition "status"   ✖

Name   status   *
            ☐ Apply for all instances

Type   Enumeration ▾
            open   ▬ ✛
            rejected   ▬ ✛
            solved   ▬ ✛
            unclear   ▬ ✛

Multiplicity   Exactly one value ▾

No draft saved yet.

Figure 4: Definition for the attribute *status* with enumeration values {open, rejected, …} and multiplicity *Exactly one value*.

3.5 Conceptual Modeling with Hybrid Wikis
Summarizing, Hybrid Wikis provide two methods supporting conceptual modeling:
- bottom-up, i.e., by the management and analysis of the wiki pages with its (type) tags and attributes
- top-down, i.e., by means of the definition of types, attributes, and consistency rules for attributes and relations

Hybrid Wikis provide the following modeling concepts:
- a (hybrid) wiki conforms to a model, defines a namespace and represents a container for the particular model elements (wiki pages)
- a wiki page represents a particular model element with URL (Uniform Resource Locator) and content (hypertext, file attachments, attributes)
- a hyperlink conforms to a model reference
- an attribute on a wiki page consists of a label (the name of the attribute) and an ordered list of attribute values
- an attribute value is either a link or a string value
- a tag classifies a wiki page, type tags (0..n) specify the type of a page, i.e., the type of the model element
- an attribute definition refers to a type tag and consists of a label and a definition of consistency rules for the number of admissible attribute values (e.g., at least one), value ranges (e.g., numbers, boolean, date, …) and valid hyperlinks (e.g., only links to wiki pages of a specific type)
- it is possible to search for incompatibilities between attribute values and attribute definitions. Furthermore these incompatibilities are marked as invalid when displaying the wiki pages.

3.6 Communication and Collaboration with Hybrid Wikis
In addition to the modeling capabilities as introduced in Section "3.5 Conceptual Modeling with Hybrid Wikis" Hybrid Wikis benefit from the services provided by the integrated Enterprise 2.0 platform Tricia. In the following we explain the interplay between these built-in services and the structuring services provided by the Hybrid Wiki extension.

In Tricia it is possible to specify extended views on the underlying data model by means of integrated search capabilities. Thereby search results can be sorted, filtered and saved. Saved queries can be embedded in the wiki page content in order to show current search hits when displaying the page. In

Hybrid Wikis users are able to define custom queries with regards to structured content, e.g., a specific attribute value.

In Tricia all changes to content objects (e.g., wiki pages) are recorded. Thereby it is traced which user has changed which content object at which time and for what reason. In Hybrid Wikis changes to structured elements are also traced. Thus, the corresponding editor (author) can directly be contacted in case of uncertainty regarding a particular change of the structure.

Furthermore in Tricia users can register to receive a notification (e.g., via email) in case of changes to a certain element (e.g., wiki page). Now, in Hybrid Wikis it is possible that users can register for changes on structured elements (e.g., changes to a specific attribute).

Tricia provides a flexible access control concept. Users can define read and write access for all content objects (e.g., wiki pages). In Hybrid Wiki this access control mechanisms also apply for attributes.

3.7 Closing the Gap Between Structured Data and Unstructured Content Within Enterprise Knowledge Management

Beside integrated Enterprise 2.0 platforms in companies a wide variety of other application is used also containing business critical data and information. These data and information are often scattered across the individual application repositories respectively data storages and are available in a structured or unstructured form (c.f., Figure 5). Examples of applications providing highly structured data are Enterprise Resource Planning (ERP), Business Intelligence (BI), and Product Lifecycle Management (PLM) systems. Unstructured information is given by applications providing objects like emails, images, and textual documents, e.g., MS Outlook, MS Word.

Since structured data as well as unstructured content may contain business critical information both kinds must be accessible to knowledge management solutions. In particular it is one of the most important technological challenges of knowledge management systems to create connections between and contextualization on these information and data in order to provide a holistic view on business critical knowledge.

Hybrid Wikis contribute to close the gap between structured and unstructured information within knowledge management solutions by providing a single user interface to create, access, search, and share business critical knowledge for these both information types. Furthermore linking within unstructured content as well as adding tags and comments allows a smooth transformation of raw unstructured data into rich understandable knowledge.

With Hybrid Wikis it is possible to describe all types of structured data by using attributes and type tags respectively attribute definitions and type tag definitions, relations by means of hyperlinks respectively (link) constraints.

Unstructured content can also be described and connected with attributes and types provided by highly structured data.

The integrated search capabilities in Hybrid Wikis enable users to create custom and uniform views on structured and unstructured information according their needs and expectations by providing full-text search mechanisms as well as features to search for tags, attributes, and other relevant facets, which can be combined flexibly. Furthermore users can easily generate hierarchical menus, tables, lists, and diagrams for information access supporting the requirements and the context of their daily business.

Since today's business environments are characterized by dynamic and rapid changes, it is often required to adapt the models underlying the knowledge management solutions, e.g., by providing additional attributes or concepts. The modeling concepts of Hybrid Wikis support this requirement by providing flexible modeling capabilities (type tags, attributes, type tag definitions, attribute definitions) and thereby allow knowledge workers to adapt the models of these solutions themselves according to the changing business demands without any need for expensive IT projects.
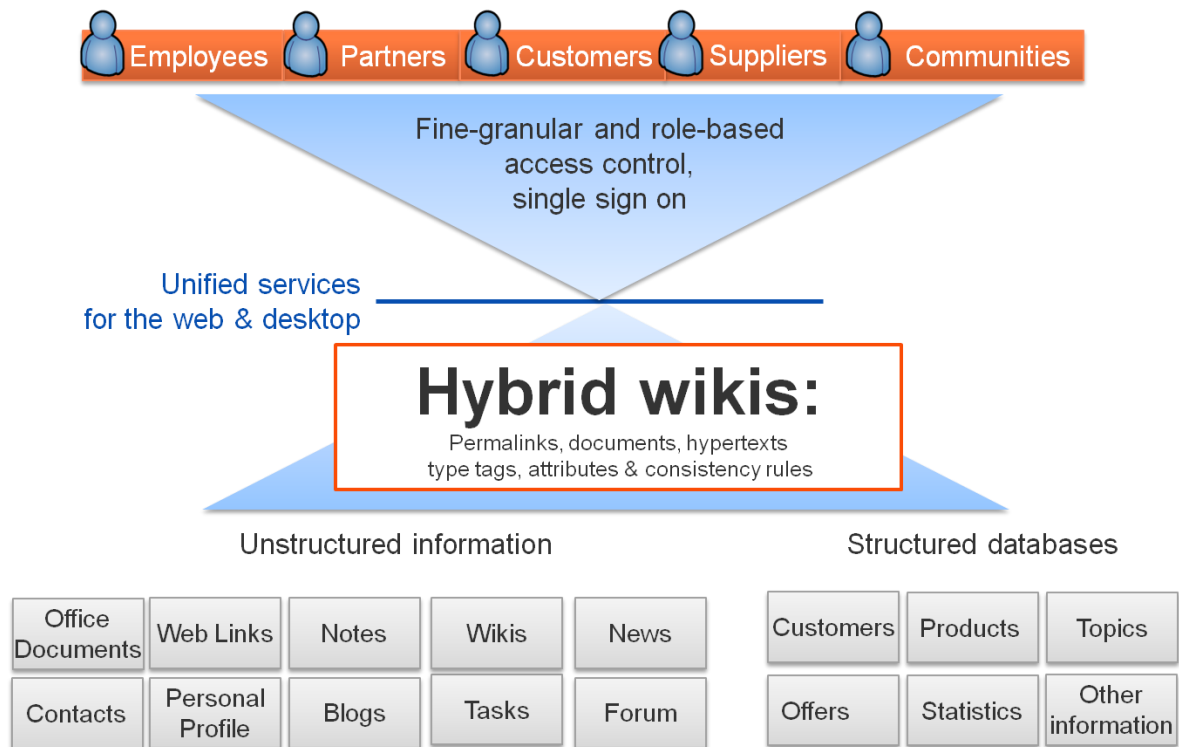
Figure 5: Closing the gap between structured data and unstructured content

4. Conclusion

In this paper, we first gave an overview of the ideas given by the Wikipedia project and Enterprise 2.0 applications. Than we illustrated the idea of Hybrid Wikis and the particular elements facilitating the structuring of content, namely type tags, attributes, type tag definitions, attribute definitions and (soft) constraints. We illustrated the application of Hybrid Wikis by providing screenshots taken from a software engineering company in which Hybrid Wikis are used as an issue tracker. Furthermore we sketched how conceptual modeling is enabled by Hybrid Wikis from two perspectives, top-down and bottom-up, and described how these structuring capabilities are integrated in the built-in services provided by the commercial Enterprise 2.0 platform Tricia.

Finally we explained how Hybrid Wikis provide a holistic view on information within the enterprise scattered across different application repositories by integrating information from both kind of data sources structured and unstructured. Thereby Hybrid Wikis contribute to bridge the gap between structured and unstructured information which is one of the key challenges in knowledge management endeavors.

References

Büchner, Thomas and Matthes, Florian and Neubert, Christian. "Data Model Driven Implementation of Web Cooperation Systems with Tricia." Objects and Databases, Lecture Notes in Computer Science, 2010: 70-84.

Chang, S. "Are they willing to contribute? Prosumer characteristics among the Australian youth." Digital Natives in Australia and Korea. Conference at the University of Melbourne, 2006.

Enterprise 2.0 Fallstudien - Aus Erfahrung lernen Webseite. 2011. http://www.e20cases.org/fallstudien (accessed 5th of January 2011).

Golder, S. and Huberman, B. A. "The Structure of Collaborative Tagging Systems." Information Dynamics Lab: HP Labs. Palo Alto, USA: http://arxiv.org/abs/cs.DL/0508082, 2005.

Haake, Anja and Lukosch, Stephan and Schümmer, Till. "Wiki-templates: adding structure support to wikis on demand." WikiSym '05: Proceedings of the 2005 international symposium on Wikis, 2005: 41--51.

Leuf, B Cunningham and Leuf Cunningham W and. The Wiki Way. Quick Collaboration on the Web. Reading, MA, Addison-Wesley, 2001.

McAfee, A. P. "Enterprise 2.0: The Dawn of Emergent Collaboration." In MIT Sloan Management Review, 47(3), pp. 21-28. 2005.

Microsoft SharePoint 2010 Webseite. 2010. http://sharepoint.microsoft.com (accessed 4th of January 2011).

O'Reilly, Tim. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. O'Reilly Media, 2008.

Resource Description Framework (RDF). 2011. http://www.w3.org/RDF/ (accessed 5th of May 2011).

Semantic MediaWiki Website. 2006. http://semantic-mediawiki.org/wiki/Semantic_MediaWiki (accessed 4th of January 2011).

Surowiecki, J. "The Wisdom of Crowds." In Random House. New York, USA, 2004.

T. Büchner, F. Matthes, and C. Neubert. "A concept and service based analysis of commercial and open." International Conference on Knowledge Management and Information Sharing. Madeira, Portugal, 2009. 37–45.

Tricia - Hybrid Wikis for Collaboration and Information Management Website. 2010. http://www.infoasset.de (accessed 3rd of January 2011).

Wiki markup. 2011. http://en.wikipedia.org/wiki/Help:Wiki_markup (accessed 5th of June 2011).