Towards Wikis as Semantic Hypermedia

Robert Tolksdorf
Free University of Berlin
Institute of Computer Science
Networked Information Systems
Takustr. 9, 14195 Berlin, Germany

tolk@inf.fu-berlin.de

Elena Paslaru Bontas Simperl
Free University of Berlin
Institute of Computer Science
Networked Information Systems
Takustr. 9, 14195 Berlin, Germany

paslaru@inf.fu-berlin.de

ABSTRACT

Similarly to the Web Wikis have *advanced* from initially simple ad-hoc solutions to highly popular systems of widespread use. This evolution is reflected by the impressive number of Wiki engines available and by the numerous settings and disciplines they have found applicability to in the last decade. In conjunction to these rapid advances the question on the fundamental principles underlying the design and the architecture of Wiki technologies becomes inevitable for their *systematic* further development and their *long-lasting success* at public, private and corporate level. This paper aims at be part of this endeavor; building upon the natural relationship between Wikis and hypermedia, we examine to which extent the current state of the art in the field (complemented by results achieved in adjacent communities such as the World Wide Web and the Semantic Web) fulfills the requirements of modern hypermedia systems. As a conclusion of the study we outline further directions of research and development which are expected to contribute to the realization of this vision.

Categories and Subject Descriptors

H.5.4 [**Hypertext/Hypermedia**]: Architectures; H.3.5 [**Online Information Services**]: Web-based services; I.2.4 [**Knowledge Representation Formalisms and Methods**]: Semantic networks

General Terms

Design

Keywords

Wikis, Semantic Web, Semantic Wikis, Hypermedia

1. INTRODUCTION

Similarly to the Web Wikis have *advanced* from initially simple ad-hoc solutions to highly popular systems of widespread

use. This evolution is reflected by the impressive number of Wiki engines available and by the numerous settings and disciplines they have found applicability to in the last decade.[1]

In conjunction to these rapid advances the question on the fundamental principles underlying the design and the architecture of Wiki technologies becomes inevitable for their *systematic further development* and their long-lasting success at public, private and corporate level. A careful analysis of these issues helps the Wiki community in clarifying and understanding the core features of Wiki systems, in designing and building extensions to them and in identifying new areas of research, development and deployment.

The basic principles of the design and architecture of the Web have been made explicit a posteriori [13, 18]. This clarification has constituted a fundamental step towards the take-up of Web-based technologies and applications beyond the boundaries of dispersed communities of practice or individual users. Due to the natural parallels between the Web and Wikis—both being in essence hypermedia systems that have grown out of practice rather than from systematic research initiatives—we believe such an endeavor is likely to have the same impact on the future of Wikis and Wiki-driven applications.

As a starting point for our investigations we build upon results established in sixty years of research in the field of hypermedia and hypertext. The basic notions underlying these systems provide us with an exhaustive, well-proven benchmark to evaluate the general-purpose usability of current Wiki solutions and to outline further directions of research, development and deployment. Of particular interest to us are *semantic technologies* that can extend both the Web and Wikis with explicit means to represent and process information not only at textual, but also at *conceptual* level. The availability of machine-understandable domain knowledge and of means to inject semantics-awareness to the traditional Web infrastructure is expected to enable or enhance the typical functionality provided by both technologies, such as *navigation*, *personalization* or *search*.

2. FROM HYPERMEDIA TO THE SEMANTIC WIKI WIKI WEB

The original ideas on *hypermedia* date back to the middle of the past century and are centered on the notion of infor-

---

[1]According to a recent statistics on the basis of `http://c2.com/cgi/wiki?WikiEngines` we counted over 200 implementations as due to February 2006.

mation items that are connected by links. Together these form a graph that can be traversed during the non-linear consumption of the linked items. The set of design options for such a system are huge (eg. unidirectional vs. bidirectional links, links with only one or with multiple destination anchors, typed nodes and links vs. plain items etc.) and have been explored in numerous systems [22].

The success and visibility of the *Web* has promoted a minimal architecture which taped by no means the full potential of hypermedia technologies. As a prominent example links on the Web are un-typed and unidirectional. Recently, various systems use the Web as a technical starting point for building more richer and usable hypermedia systems [6, 29].

Wikis are Web-based applications that provide content authoring and management functionality in a much simpler manner than its Web counterpart [20]. They simplify the hypertext generation task by offering a restricted syntax for information markup and browser-integrated editing capabilities. Every Wiki instance is a set of dynamically generated HTML pages—its content consists of an overlay network of Wiki articles on the underlying Web platform with its network of HTML pages. Links between articles are handled at the application layer, which resorts to an explicit data model of hypertext. In contrast to the traditional Web, however, Wiki systems store linking information persistently in a database, thus providing link bi-directionality.

The *Semantic Web* overlays a typed conceptual graph on the conventional Web contents [5]. It shifts links and content to this additional level, allowing richer navigation, search and personalization facilities with the help of automatic inferences [11, 14, 27]. The Semantic Web is based on conceptual structures (termed *ontologies*), standard Web-compatible knowledge representation languages (RDF(S)[16, 8], OWL[23] and, newly, SWRL[17]) and theoretical logics that guide their use.

Even more recently the combination of Wikis and Semantic Web has gained attention, leading to a set of so-called *Semantic Wikis* proposals. While applying semantic technologies to conventional Web (or Wiki) information leads to a Web that can be better read and processed—by using inferencing for higher precision and recall, Semantic Wikis also aim to provide a low-entry barrier means to boost the large scale generation of semantically rich information [3, 28, 25]. Being based on both an application level manageable hypertext model and on formal conceptual structures they are considered by many members of the two communities as a promising alternative to current approaches for collaboratively creating and managing knowledge.

Figure 1 illustrates our view on the named technologies along two axis: the Web and the Semantic Web are means to *represent* content and knowledge and Wikis and their semantic counterparts are means *to create and manage* them.

Within this framework we study the four mentioned technologies against an established set of requirements on modern hypermedia systems. We first review these requirements as postulated in recent hypermedia literature. Then we analyze to which extent each of these technologies already fulfills the introduced requirements, identify missing functionality, and suggest ways to overcome these drawbacks. As a result we are able to present a set of directions for further research in Semantic Wiki systems necessary to bring them close to the original vision of hypermedia.

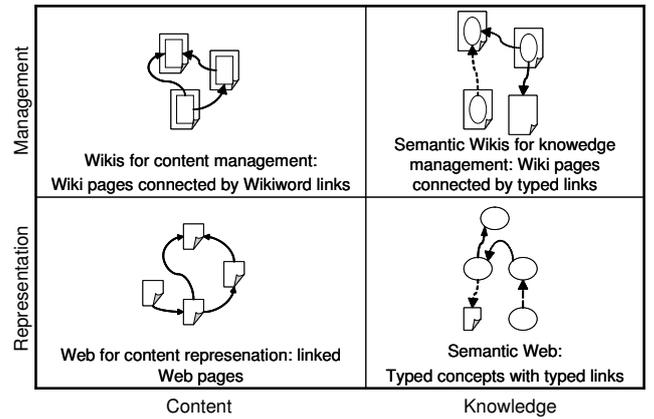The remainder of this paper is organized as follows. Sec-

Figure 1: Hypermedia Structures under Consideration

tion 3 introduces the requirements which should be fulfilled by the analyzed technologies in order for them to provide real added value to arbitrary application scenarios at large scale. These act as an evaluation benchmark for studying the current state of the art in Section 4. We discuss our findings in Section 5 and conclude with a summary and outlook of this work in Section 6.

3. REQUIREMENTS ANALYSIS

The requirements analysis is illustrated in terms of a simple application scenario. We first introduce the dimensions of this scenario, which is situated in the Internet service provision sector, and then explain how hypermedia systems (such as Web portals, Wikis and their semantic extensions) could contribute to its realization.

3.1 Application Scenario

In order to improve the quality of its customer-related services an Internet service provider company plans the implementation of a Web-based technical support center assisting business and private customers in effectively using specific ICT products. The prospected hypermedia system will deliver interested parties information about the available range of products, facilitate the selection and purchase of customized product packages and provide existing customers with technical consultancy according to their specific needs. Besides potential and existing clients the system is expected to provide support to particular company departments which publish and use Web and intranet information in order to exercise their everyday business. With this respect we exemplarily consider marketing assistants generating and managing information on products and services and IT personnel required to aid individuals in issues related to installing, using and updating the delivered products.

The way the system will satisfy the needs of each of the aforementioned target user groups will vary w.r.t. the amount and the type of information delivered, the policies concerning the access to this information and the presentation layout. General information may include descriptions of common product packages and of the ways they can be used in standard circumstances and is made available to the public on the Web. Existing customers may receive personalized promotional information or guidelines for optimizing

the utilization of currently purchased products (e.g. information about new releases or upgrades). Moreover, company personnel should additionally have access to details regarding customer management, sales, pricing and technical assistance and internal working procedures.

The system will need to manage a variety of *interrelated* information sources, from product descriptions, user manuals and installation howto's to best practices and guidelines, FAQs or online forms, many of them dedicated to specific target audiences. The content will be *jointly* produced by different parties. Marketing personnel will be principally responsible for creating product descriptions. IT specialists will be involved in the generation of user manuals and installation instructions. Furthermore, the system may allow *clients* to independently contribute to the creation of best practices and guidelines, which will be revised by the system operator, and to share their product experiences with other parties. The content authors will need means to collaboratively create, adapt and update each information item according to its prospected usage contexts or in response to new situations. The system will therefore provide features for supporting collaboration between team members, as well as multiple views and versions of the managed information resources.

A last, but important feature of the prospected technical support center will be the integration of external systems addressing topics which are relevant for the company's business. One can imagine third party discussions forums concerned with the Internet service provision market or enterprize portals of external suppliers such as hardware manufacturers or software developers which act as infrastructure for the delivered Internet services.

This functional synopsis corresponds to a series of system components as follows:

authoring component: The system should provide means for creating and managing new information sources. In particular this requires means for:

> **information organization:** the huge body of heterogeneous information sources need to be organized and interrelated in a meaningful way.

> **collaborative authoring:** as information is managed collaboratively, multiple users need system support for creating and editing a shared version of the same information resource.

> **versioning and updates:** due to the inherent nature of this distributed information management scenario the system is required to explicitly provide an answer to the question of consistently and flexibly managing a network of dynamically changing information resources.

usage component: Given a network of meaningfully interrelated information sources the aforementioned user communities need means to access this repository:

> **information retrieval:** in order to handle the available information the system should include search and navigation features.

> **personalization and context:** the way information is accessed can be optimized by taking into consideration the personal profile of the users and

the context of the activities currently being performed.

security and privacy component: As multiple user groups access the same information repository it is essential to control and monitor this procedure with the help of appropriate policies and security mechanisms.

integration component: The system needs means to consistently relate to external information sources in order to improve the quality of its information services and to round out the functionality of the overall system.

In the remaining of this section we address the implementation of the Internet service provider scenario in terms of features of hypermedia system and beyond.

3.2 Hypermedia-oriented Requirements

Table 1 depicts a compilation of the most important features of modern hypermedia systems, which are expected to contribute to the realization of the aforementioned scenario. Their relevance to the functional architecture introduced in the previous section is grounded in hypermedia/hypertext literature [6, 7, 29] and briefly outlined in the next subsections.

No.	Hypermedia feature	System functionality
F1	Types and attributes for nodes and links	information organization information retrieval versioning and updates personalization and context
F2	Transclusions	authoring versioning and updates personalization
F3	Annotations and public/private links	information organization personalization and context collaborative authoring security and privacy
F4	Computed personalized links	information organization link updates personalization and context
F5	Separation between nodes and links	personalization and context versioning and updates
F6	Global and local overviews	information organization information retrieval personalization and context
F7	Backtracking and history-based navigation	authoring usage
F8	Trails and guided tours	information organization information retrieval personalization and context
F9	Edit-capable browsers	collaborative authoring

Table 1: Features of Modern Hypermedia Systems and Their Usage in the Application Scenario

Complementarily to these "classical" hypermedia features we mention the need for semantic technologies to enhance automatic link computation, advanced information retrieval

and organization as well as content quality assurance facilities (cf. Section 4.2 below). Further on, as illustrated in Table 1, hypermedia can not provide fully-fledged solutions for security, privacy or application integration, though the implementation of these components might be in relationship with certain hypermedia design decisions.

In the following we give a short explanation of each of these features outlining their role in the previously introduced application scenario.

3.2.1 (F1) Types and Attributes for Nodes and Links

Categorizing nodes and links provides an effective means to organize and retrieve information sources in a system or an enterprize. Associating pre-defined types to the former contributes to the enhancement of the retrieval and presentation functionality of the system, which can be optimized according to this classification. By typing links content authors are able to specify the meaning between the current and the destination node, a pre-requisite for more intuitive navigation facilities [4]. Attributing these items reinforces these effects. Additional domain- or task-related information such as keywords, time stamps, authors, versions, target audience in relation to nodes or links form the basis for the realization of context-specific services.

In terms of the Internet service provider scenario, a meaningful node classification might be related to the subject described in the corresponding information source (e.g. a product such as `DSL-2000-flat` which is marketed by the company) and the type of information material (e.g. `howto`, `brochure`, `flyer`, `product overview`). The link types are defined accordingly. Domain-specific links might connect subject categories (e.g. products and associated software); further on, one can imagine domain-independent links such as `explains`, `details`, `isPreviousVersionOf` which describe meta-relationships between different kinds of information materials. On the basis of this differentiation the system could provide structure-based query and browse functionality. When an employee is searching for a particular resource, the retrieval component might for instance automatically include to the search space all additional nodes connected through the `isPreviousVersionOf` link with the nodes fulfilling the query. Customers could receive personalized information material, which is `similar` to the products already purchased, while the similarity measures would be implemented in terms of link types and attributes.

3.2.2 (F2) Transclusions

Transclusions, a "pointer-like" mechanism for having the same contents co-exist in multiple places, were one of the key features proposed by hypertext initiators to solve redundancy and update problems [21].

Transclusions provide content authors with a flexible means to create and maintain documents with overlapping contents. If systematically implemented they play a key role for realizing cross-enterprise and cross-application interoperability. Turning back to our application scenario, one can imagine that much of the promotional material produced by the marketing department does contain the same basic contents, but in a different order, layout or level of detail. Installation guides for example contain a part which captures details about the Internet access provision and one related to the IT infrastructure required (software programs, hard-

ware etc.). These in turn might be associated to various charges and fees, and to variants and releases, respectively. Transclusions would allow marketing personnel to easily assemble these items to new documents and to maintain them in a consistent manner.

3.2.3 (F3) Annotations and Public/Private Links

From the early days of hypermedia annotations to published contents have been understood as a basic means to enable collaborative tasks such as content authoring [6]. A complementary issue is whether and to which extent to allow users to directly customize the data, the associated layout or the available links.

Creating views and restricting the rights to interact with the hypermedia contents requires customization based both on user roles and access permissions. Access policies (read, write, delete) might be applied at document, but also at link and annotation level. For certain documents or document types the system could foresee link or annotation authoring while restricting the write access to the underlying content. Moreover, link types and their visibility (public vs. private) might refine the access policies applied.

The Internet service provider may allow users to contribute to the generation of best practices, guidelines and howto's, which are shared among the customer community. However, they will restrict access to internal procedures applying to (particular groups) of employees. This differentiation might be further refined at the level of links and annotations: annotations might be deleted by their author and by system operators, while link management might be differentiated to the type of connected documents: if, for instance, the links relate to an area with a un-restricted access to public, they might also decide upon the form and the content of the links.

3.2.4 (F4) Computed Personalized Links

Link computation is a powerful feature of modern hypermedia-based information management systems. They are intended to ease users the creation of customized links between information sources in response to the rapid evolution of the hypermedia contents.[2]

A marketing agent creating customer-specific promotional offers may wish to automatically create a link between products and associated software and hardware requirements. A technical assistant might use the same facility to connect best practices and FAQs to product features. If these links had been created manually, the users would have to create or maintain them every time changes occur in the corresponding information.

The automatical generation of (personalized) links implies however the availability of mechanisms to separate this type of information from plain content (see below).

3.2.5 (F5) Separation between Nodes and Links

Separating information sources from the links interconnecting them enables the realization of a wide range of hypermedia features related to personalization and other context-

---

[2]As stated at `http://stats.Wikimedia.org/DE/TablesWikipediaEN.htm` the English Wikipedia, one of the most successful contemporary hypermedia system, contained 19.3 million internal links to 922.000 articles as of November 2005. This is a rate of approximately 21 out- and in-links per article.

and structure-based services: different link sets over the same content for different target user groups or tasks, private and personalized links, and finally bidirectional links [29].

For the Internet service provider scenario storing links externally is a pre-requisite for the implementation of various accountability services. Links to older versions of the products on offer might be preserved in the system for historical or legal reasons. Links to external information sources describing software and hardware infrastructure supplied by business partners should be managed independently of the present availability (or unavailability) of Internet service provision offers building upon them. If the range of ICT products delivered by our enterprize at a moment in time does not necessitate a particular software or hardware component, the link to the documents describing these components should be preserved within the system in order to be able to refer to these information items when new offers become available.

3.2.6 (F6) Global and Local Overviews

Overviews are implemented in many proprietary hypermedia solutions as a means to aid navigation. The difference between global and local ones is directly related to the level of granularity of the presented information. The underlying generation mechanism highly depends on the features provided by the system at node and link level: from examining the nodes for embedded links up to sophisticated crawling based on user preferences on links and documents [22].

In terms of the technical support center such overviews are motivated by the large amount of highly interconnected information items available. An IT assistant guiding a customer w.r.t. installation questions may be interested in finding out the local and global context of the corresponding product. Alternative product variants, similar products or classes of products might be associated to similar installation characteristics. In consequence the best practices and guidelines available might be valid across these categories and offer valuable hints for solving new technical problems. This kind of information is hard to assimilate when the technical assistant has access to isolated content resources without being aware of the position of these sources in the overall context. Overviews are in this context the natural means to a priori aggregate and materialize heterogeneous information sources on the basis of application-specific relevance criteria.

3.2.7 (F7) Backtracking and History-based Navigation

These are features of hypermedia systems which allow users to browse the underlying information collection and to return to previously visited nodes without any particular knowledge on the exact structure of the hypertext. The range of these navigation facilities can be of arbitrary complexity, again depending on the availability of basic link and nodes features like types, attributes and annotations. Building upon this additional information a hypermedia system might implement various parameter-based navigation support (e.g. based on conditions, on tasks, on personal profiles) complementarily to the conventional chronological navigation history.

In terms of the Internet service provider scenario these advanced services permit content authors and users to more

comfortably access hypermedia information without getting lost in the hyperspace of product descriptions, howto's, installation kits, FAQs and user discussion forums [24].

3.2.8 (F8) Trails and Guided Tours

Trails denominate link sequences created on user or application-specific criteria [9]. Guided tours are a particular type of trails in which users can unidirectionally navigate the information sources along the path in pre-defined ways [15].

We can imagine various utilizations for guided tours and trails in the context of the technical support system introduced above. Trails generally allow users to optimize their navigation sequences across the comprehensive network of heterogeneous multi-versioned information sources. Guided tours might be created for specific customers, for users with different levels of IT expertise or for the public interested parties in order to ease the comprehensiveness of the product descriptions. Company personnel can be trained in various topics related to their everyday business, while external parties might be given a brief tutorial on the way they can interact with the system.

3.2.9 (F9) Edit-capable Browsers

Many of the hypermedia systems emerged in the pre-Web era considered the integration of editing and presentation as one of the key features for promoting wide scale participation and enabling collaborative content authoring and management [29].

As mentioned in the previous section, the prospected system relies on the voluntary participation of its customers to generate and maintain reliable and useful installation support materials on the basis of their everyday experiences with the purchased products. This implies in particular a low barrier of entry for utilizing the system, a pre-requisite which is considered by many hypermedia research as being optimally implemented at browser level (cf. [29] for a discussion on this topic).[3]

4. TECHNOLOGY EVALUATION

On the basis of the feature set described in the previous section we examine the state of the art in aforementioend four technological fields: the World Wide Web, Wikis and their semantic counterparts, the Semantic Web and Semantic Wikis, respectively.

4.1 The Web as Hypermedia Representation of Content

The expressivity of the Web as a hypermedia system is largely dominated by the concepts embodied in HTML (as currently defined by XHTML 1.1 [1]) and the ones foreseen in HTTP to process Web-based interaction.

The Web is commonly seen as a simplified version of a fully-fledged hypermedia system [6, 29], the most important restriction being probably the absence of a usable concept of *typing* both links and pages.

The `<meta>` element can be interpreted as a substitute for this missing typing functionality, since it can express that a certain metadata entry has a certain value following a certain scheme. While this does in fact allow one to physically

---

[3]Edit-capable browsers, such as W3C's Amaya, differ from conventional browsers provided with compatible content authoring components such as FrontPage for Internet Explorer or Composer for Mozilla.

encode for example, that a page is about a certain topic following e.g. the ACM classification scheme, there is no defined way how to *interprete* that information automatically. This facility is expected to be implemented at client level. Therefore, HTML does not contain any dedicated means to type pages which is integral part of Web technologies.

The `<link>` element can interrelate documents by selecting from a set of predefined link types such as `Chapter`, `Help` and `Next`. The set of links is oriented towards document structure and aims at supporting agents like browsers (preloading a potential subsequent Web page, using a stylesheet) or search engines (marking the start of a document collection). There is no mechanism in place for a systematic extension of this typing mechanism. Further on, while the `<link>` element allows to express the direction of the relation (i.e. with the help of the `rel` and `rev` attributes), this is definitely not equivalent to bidirectional links support. Some browsers do take advantage of the structural information by display `Next` or `Glossary` buttons in the navigation interface.

The attributes `rel` and `rev` also apply to the `<a>` element, which is otherwise an un-typed unidirectional link. While `<link>` relates complete documents, the granularity of typed `<a>`-links remains unclear. For example, one could state that we are at a subsection of a chapter (`<a href="http://foo.bar/section4.3.html" rev="Subsection">`); however this does not clarify the question on the scope of this "typing". We would have to enclose the complete subsection with the `<a>` element but this would generate a huge anchor in the browser that would lead to the hierarchical enclosing section(-document). Since typing is absent or broking in HTML, we cannot expect to express anything about link attributes and structures.

A "taste" of transclusions is present in HTML in tags like `<img>`, `<object>`, or `<applet>`. However, these are not part of the HTML standard and implementing them implies additional modules within Web servers, which are not seamlessly integrable [19].

Annotation, public vs. private links were never part of the HTML/HTTP technology. They were in fact included in the early Mosaic browsers that could contact an "annotation server" where one could leave comments on a page. Today, this idea is revitalized in a simplified form by services such as de.licio.us[4] or to a larger extent in Weblogs.

External link databases, absent from the early days of the Web, resulted in alternative Web architectures, the most notably of which being probably Hyper-G [2]. This system was built on a set of link and document servers with links being first class objects. Thereby, they were bidirectional and Hyper-G featured global and local overviews as integral parts of its browsers. As a side effect it enabled backtracking and history-based navigation. While Hyper-G had a Web interface it comprised a completely different architecture and never had a substantial impact. On the Web document content and link anchors are not separable.

Edit-capable browsers are meanwhile available, most prominently the Amaya browser by W3C. These could in fact work on any sites by HTTP-methods like PUT or DELETE. However, the currently employed Web infrastructure disables the usage of these methods for good security reasons.

These considerations are summarized in Table 2.

---

[4]http://del.ico.us

FEATURE No.	IMPLEMENTATION	USAGE
F1	minimal in HTML	sporadic use
F2	minimal in HTML	average use
F3	Weblogs folksonomies	wide use average use
F4	-	-
F5	-	-
F6	-	-
F7	minimal in browsers	sporadic use
F8	Web portals	sporadic use
F9	some browsers available	sporadic use

Table 2: The Web as a Hypermedia Representation of Content

4.2 The Semantic Web as Hypermedia Representation of Knowledge

The Semantic Web focuses technologically on a set of XML-based languages, used to annotate plain HTML content with `(subject, predicate, object)` statements, whose meaning might be constrained by conceptual schemes (i.e. ontologies) and processed by machines.

The Semantic Web[5] envisions a distributed network of machine-understandable knowledge. As it is built upon the existing Web infrastructure it inherits the Web architectural model, which has been formally described as REST (Representational State Transfer)[13]. The fundamental principle of the REST architecture is that resources are *stateless* and *published* based on a *global and persistent URI*. The Semantic Web extends this world of URIs from addressable resources to any abstract concepts represented within a computer system. By means of ontologies one can formally specify resource types, define attributes and properties of these resources and annotate them with additional information. In this way, the Semantic Web approaches several fundamental limitations of the traditional Web related to the enhancement of contents and links semantics (cf. Table 1, features F1 to F3).

As contents are represented (or annotated) using formally defined conceptual structures, the Semantic Web can apply reasoning services to automatically derive links between semantically related information sources.

Personalization can be implemented on the top of semantic technologies. Web resources may be annotated with concepts defined in arbitrary schemes online available. If personal profiles are available in this form links and nodes can uniquely refer to particular user preferences captured by the profile [11, 14].

For the same reasons the Semantic Web promotes the separation between contents and links. Link and node types are defined in a machine-understandable conceptual model (i.e. an ontology). Hypertext nodes are represented as instances or individuals of a particular type and are interconnected by links externally referencing the link types defined in the model. Which links are allowed between two nodes (link computation) is clearly specified in the ontology by means of domain and range definitions or logical axioms. The OWL language additionally supports the definition of symmetric and inverse properties; by representing a link as an OWL property and associating it with a symmetry constraint or

an inverse property (i.e. a reverse link) we can easily define bidirectional links.

Trails can be built upon transitivity properties in OWL or SWRL rules. The latter are a useful means to declaratively represent parametric conditions for customize navigation or to define the allowed trails within a guided tour.

Nevertheless collaborative aspects have been marginally explored in the Semantic Web community yet. Research efforts are concentrating for the moment on issues related to the representation of knowledge and its usage within scalable reasoning services. On the other side, recent advances in these areas provide a feasible basis for the implementation of information retrieval and integration components which are out of the scope of hypermedia [12]. Ontologies are widely acknowledged as the key enabler for the realization of semantic search heuristics based on content-narrow logical inferences (as opposed to keyword- or statistics-based techniques which do not take into account domain knowledge). In conjunction with expressive representation languages such as OWL they can be also used to define equivalence or mapping relations between heterogeneous information spaces and their local categorization structures, thus facilitating application interoperability.

An overview of the evaluation of this technology w.r.t. the hypermedia features introduced in Section 3.2 is depicted in Table 3.

FEATURE No.	IMPLEMENTATION	USAGE
F1	ontologies to define types and attributes	sporadic use
F2	URIs, semantic node structure	sporadic use
F3	ontologies to describe personal profiles and annotation types	sporadic use
F4	OWL and rules reasoning	sporadic use
F5	Semantic Web architecture	sporadic use
F6	OWL and rules reasoning	sporadic use
F7	ontologies to declare parametric navigation	sporadic use
F8	OWL and rules reasoning	sporadic use
F9	the Amaya browser	sporadic use

Table 3: The Semantic Web as a Hypermedia Representation of Knowledge

4.3 Wikis as Hypermedia-based Content Management Systems

Wikis [20] reify Web pages and links at application level (cf. Figure 1). A concept is described on a Wiki page which is in turn rendered as HTML. The page can be edited using a simplified markup language and the editor is offered as an HTML-form. Wiki pages are linked directly by WikiWords.

Since Wikis treat pages as first class objects—and usually store them in databases—they could extend their data model to include type information for pages. However, this is not commonly the case. Wikimedia introduced the notion of "category" [31] to attribute a Wiki page to be of a certain kind. A declaration `[[Category:Howto]]` classifies the page into the category denoted by the tag `Howto`. These categories can be queried on specialized pages computing category lists or category-centered page clusters. A page can be ordered to multiple categories and one can define sub-categories by editing a category page and declaring its super-category. For links, Wikimedia introduces a minimal structural typing [32] for referring to a parent-, sibling- or sub-page with a link as in `[[../Sibling]]`.

This set of categories focusses on simple structural aspects and user interface issues and can not be interpreted as a fully fledged typing mechanism for hypermedia objects. The sub-category relation does not posses a formal semantics (e.g. does a page attributed to be of a category also belong to a possible super-category?) and there is no way of expressing queries that would exploit such a structure automatically. Moreover, there are no means to express relations amongst types aside from the sub-category relation.

The purely document structure-oriented typing of links is rudimentary and at the same level of expressivity as the referred characteristics of the `<link>` element in HTML. The link "type" information is encoded into the link itself, therefore it is not possible to declare further attributes on links.

If one considers information on a Wikipage to be typed by its WikiWord, Wikimedia lacks an extensible mechanism to dissolve name clashes. There is a namespace mechanism present, however, it is not extensible within the Wiki framework[5].

Wikis today can display solely very simple global and local overviews. This might be due to the initial restrictions of HTML technologies relevant for the time the first Wikis were developed. However, this drawback seems to be reducible to the technical issue of whether one could introduce some display with Java applets or AJAX technologies.

Wikis do not offer much personalization of content. By skins there is a way to select from personalized representation, however, the content displayed is always the same. Further on Wikis do not keep an interaction history and therefore cannot offer backtracking and history-based navigation.

Wikis contain per definition an edit-facility for Wiki pages. This is at the core of the Wiki concept, that of collaboratively managing the content [20]. Since the complete Wiki is reified as a Web application, the edit-facility is not some external program but an HTML form which is processed at the server side. Currently, the usability of such editors is below that of more common editor applications. However, the currently developing AJAX technologies can fill that gap without tradeoffs.

The built-in editing facility is of course also the basis for annotating information within the Wiki. Either one can change the information itself, or one can introduce a new page containing annotations. MediaWiki features a discussion service with the `Talk` category which can be used as an annotation service.

Again we summarize our results in Table 4 below.

4.4 Semantic Wikis as Hypermedia Knowledge Management Systems

A multitude of approaches for combining Semantic Web and Wiki technologies are currently under development (cf., for example, [3, 25, 10, 26, 28, 30]). They all share the same goal of extending the functionality of Wiki engines with means to create and manage semantic data. This in

---

[5]One has to alter PHP code to introduce namespaces for interwiki links.

Feature No.	Implementation	Usage
F1	minimal as categories	wide use
F2	-	-
F3	discussion pages	wide use
F4	minimal personalization with skins	average use
F5	-	-
F6	simple pre-defined views	average use
F7	-	-
F8	-	-
F9	per definition	wide use

Table 4: Wikis as Hypermedia-based Content Management Systems

return provides a flexible means for typing and annotations. Additionally it enables the usage of automatic inference services, which can be applied to automatically generate new information (e.g., new links) for more sophisticated retrieval and personalization strategies (cf. Section 4.2).

These features are however not supported by any of the available implementations to a satisfactory extent. Nevertheless Semantic Wikis are considered by many researchers and developers in the field as a promising approach to (at least partially) overcome many of the limitations of the two technologies. This idea becomes clear by comparing Tables 3 and 4.

Approaches such as [26, 28] develop Wiki engines which support the generation of RDF data. However, they clearly distinguish between semantic and plain Wiki contents and their usage in Semantic Web context requires technical expertise on RDF. Acknowledging for this limitation more recent proposals focus on the minimal invasive usage of semantic technologies within Wiki systems[3, 25, 10, 30]. The Semantic MediaWiki has come up with a different approach [30]. The aim of this project is to turn the most prominent existing Wiki-based information repository—the Wikipedia encyclopedia—into a semantic Wiki. Similar to [3, 25, 10] the developers of SemanticMediaWiki identified the need for typed annotated links and articles as central for Wikipedia and extended the Wiki engine with simple means to create RDF triples in addition to plain Wiki text and to export the RDF contents. The proposals described in [3, 25, 10] refine this functionality to support the definition and usage of external ontologies as typing structure for Wiki pages, contents, annotations and links. [10] additionally provides a prototypical implementation of an ontology-based search component for the JSPWiki engine.

5. DIRECTIONS OF RESEARCH AND DEVELOPMENT

The analysis of the four technologies as regarding the extent to which they implement the core features of modern hypermedia revealed several important limitations of current Wiki solutions. While the Wiki Way definitely approaches one of the main drawbacks of the conventional Web—that of edit-capable browsers— and additionally provides feasible means to enhance collaboration and type hypermedia information (the Semantic Wikis) several important issues remain unsolved.

An essential limitation is in our opinion the design decision considering a Wiki as a *closed* world with no interaction possibilities with other Wikis or other services. At the level of the introduced application scenario and beyond, it is required to introduce *distribution* and *openness* as integral design concepts for future Wiki systems. We argue on the necessity of systematically implementing the following three features:

- having links and pages as *first class objects* to enable the design and realization of new external services,

- having a standardized *object model* and a standardized, remotely accessible *API* to allow for a distributed architecture in which these services can evolve, and

- having *processable semantic descriptions* to considerably improve the quality of such a World Wide Wiki.

Making links and pages accessible in a distributed fashion calls for a standardized object model. Such a *Wiki Object Model* (WOM) would be similar to what the Document Object Model is for the Web.[6] If properly defined this type of virtual objects can form the basis for a variety of powerful operations such as persistent storage. Necessary is an organizational framework which initiates and coordinates the activities related to the creation of such a WOM. It should be the best practice of the object models (mostly implicitly) embodied in current Wiki systems.

The WOM model should be defined in a standardized manner which already supports distribution and openness as a design concept. Following the DOM example it is likely that the CORBA-IDL is a reasonable starting point for this activity.[7] We also propose that the WOM developers nominally take into account semantic descriptions expressed in standardized Semantic Web representation languages such as RDF in order to promote the compatibility with the newly emerging projects around Semantic Wikis. Due to the inherent concept of object identity, a WOM will also provide a uniform way to access Wiki pages and links across different systems as a side-effect, for example by identity representation as a new URL scheme.

Further on, the WOM has to be accompanied by an API that offers a standard way to access the objects contained in a Wiki. If we follow the previous idea of a programming language independent definition of the WOM the API can be implemented in various languages. This includes libraries for both access to remote Wikis and for internal programmatical access within a Wiki (required by something that we would call WikiScript, which would be an execution environment within a Wiki with programs embedded in Wiki pages).

If Wiki pages and links are defined as first class objects that have a uniform structure and a uniform identity, both representable in terms of URLs, then semantic statements can be made on pages as well as on anchors and links using currently available Semantic Web technologies (cf. Section 4.2).

Recalling the hypermedia features illustrated in Table 1 we can factor out several of these concepts to *services* based on a WOM and remote access to it:

(F6) Global and local overviews would be offered as external services that can be adapted to the respective

---

[6] http://www.w3.org/DOM/
[7] http://www.omg.org/gettingstarted/omg_idl.htm

usage context when the overview is required. Via an API they have full access to the structure under consideration even if scattered over various Wikis and full freedom in generating suitable layouts. Semantic descriptions can considerably improve the quality of such visualizations, e.g. by providing a measure of semantic "relatedness" that is mapped into spatial distance in a visualization.

(F3, F4) Personalization could be implemented by sets of internal programmatic accesses to both the content of a Wiki and some external representations of a personal context or profile. Semantic technologies greatly support the expressiveness of such representations and allow for more precise decisions.

(F7) Backtracking and history-based navigation would also be external services that manage a personal database of previously accessed objects, that is pages and links. Such databases can be completely decoupled from the Wiki engine and integrated with other personal services. Semantic information can help in inferring and supporting the *intended* interaction within a Wiki instance.

Multimedia integration would be an inherent part of a WOM and its supported data types. Semantic information helps in making that support more sophisticated by supporting the integration of multimedia information.

(F4) Computed links might provide most of the immediately visible added value of the proposed structure. As explained in Section 4.2 they could be the result of Wiki internal processing and reasoning, but also subject for external services whose outputs could be seamlessly integrated into a Wiki as remote pages or anchors. With semantic information and an adequate infrastructure at hand, the quality of such *inferred* links is upvalued to a new level.

To date the Wiki community seems to be several steps away from the realization of such a World Wide Wiki. In order to fill this gap we envisage the following research and development roadmap:

- Standardize a data model part of the WOM (Wiki Object Model) and augment it with semantics.

- Standardize the API part of the WOM and augment it with semantics-aware functionality.

- Build reference implementations of WOM-based Wikis, implement APIs in multiple languages.

- Define, implement and make basic services that demonstrate added value in the areas overviews, global and personal history and inferred links based on a WOM freely available.

6. SUMMARY AND OUTLOOK

Building upon the natural relationship between Wikis and hypermedia, this paper examines to which extent the current state of the art in this field (complemented by results achieved in adjacent communities such as the World Wide Web and the Semantic Web) fulfills the requirements of modern hypermedia systems. As a conclusion of this study we are able to outline further directions of research and development which are expected to contribute to the long-lasting success of Wiki technologies at public, private and corporate level. In particular we argue for the need of a standardized semantics-driven Wiki Object Model which forms the basis for the realization of integrated Wiki solutions at Web scale. We also propose the introduction of advanced hypermedia features, whose added value is grounded in an impressive body of research in the hypertext and hypermedia areas, in terms of so-called WikiServices.

In the future we intend to continue this research endeavor with a careful estimation of the potential of newly emerging Web 2.0 concepts such as meshups as a means to enable the component-based development of trails and guided tours within Wikis. A second research direction is related to the topic of information retrieval. With this respect we are approaching search heuristics taking into account collaborative and social aspects in addition to semantics.

Acknowledgments

7. REFERENCES

[1] M. Altheim and S. McCarron. XHTML 1.1 - Module-based XHTML. W3C Recommendation, 2001. *http://www.w3.org/TR/xhtml11*.

[2] K. Andrews, F. Kappe, and H. Maurer. Serving information to the Web with Hyper-G. In *Proceedings of the Third International World-Wide Web conference on Technology, tools and applications*, pages 919–926, New York, NY, USA, 1995. Elsevier North-Holland, Inc.

[3] D. Aumüller. Semantic Authoring and Retrieval in a Wiki. In *Demo Session at the European Semantic Web Conference ESWC2005*, 2005.

[4] D. Benyon, D. Stone, and M. Woodroffe. Experience with developing multimedia courseware for the www: the need for better tools. *International Journal of Human-Computer Studies*, 47, 1997.

[5] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, 2001.

[6] M. Bieber, F. Vitali, V. Balasubramanian, and H. Oinas-Kukkonen. Fourth generation hypermedia: some missing links for the World Wide Web. *International Journal of Human-Computer Studies*, 47:31–65, 1997.

[7] M. Bieber and J. Yoo. Hypermedia: a design philosophy. *ACM Computing Surveys*, 31(4):29, 1999.

[8] D. Brickley and R. V. Guha. RDF Vocabulary Description Language 1.0: RDF Schema. *Available at http://www.w3.org/TR/rdf-schema/*, 2004.

[9] V. Bush. As we may think. *Atlantic Monthly*, 176:101–108, 1945.

[10] C. Dello, E. Paslaru Bontas, and R. Tolksdorf. Creating and using semantic content with Makna. In *Proceedings of the 1st International Workshop Wikis meet Semantics co-located with the ESWC2006 (to appear)*, 2006.

[11] P. Dolog, R. Gavriloaie, W. Nejdl, and J. Brase. Integrating adaptive hypermedia techniques and open RDF-based environments. In *Proceedings of the 12th International World Wide Web Conference WWW2003*, 2003.

[12] D. Fensel. *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*. Springer Verlag, 2001.

[13] R. T. Fielding and R. N. Taylor. Principled design of the modern Web architecture. *ACM Transactions on Internet Technology (TOIT)*, 2(2):115–150, 2002.

[14] F. Frasincar and G. Houben. Hypermedia presentation adaptation on the Semantic Web. In *Proccedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems AH2002*, 2002.

[15] F. Garzotto, L. Mainetti, and P. Paolini. Navigation in hypermedia applications: modeling and semantics. *Journal of Organizational Computing and Electronic Commerce*, 6(3):211–237, 1996.

[16] P. Hayes and B. McBride. RDF Semantics. *Available at http://www.w3.org/TR/rdf-mt/*, 2004.

[17] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, and M. Dean. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. *Available at http://www.w3.org/Submission/SWRL/*, 2004.

[18] I. Jacobs and N. Walshauthor. Architecture of the World Wide Web, Volume One. W3C Recommendation, 2004. *http://www.w3.org/TR/webarch*.

[19] J. Kolbitsch and H. Maurer. Transclusions in an HTML-Based Environment. *Journal of Computing and Information Technology*, 14, 2006. to appear.

[20] B. Leuf and W. Cunningham. *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley, 2001.

[21] T. H. Nelson. The heart of connection: hypermedia unified by transclusion. *Communications of the ACM*, 38(8):31–33, 1995.

[22] J. Nielsen. *Hypertext and hypermedia*. Academic Press Professional, Inc., San Diego, CA, USA, 1990.

[23] P. F. Patel-Schneider, P. Hayes, and I. Horrocks. OWL Web Ontology Language Semantics and Abstract Syntax. *Available at http://www.w3.org/TR/owl-absyn/*, 2004.

[24] J. Rosenberg. The structure of hypertext activity. In *Proceedings of the Hypertext96*, pages 22–30, 1996.

[25] S. Schaffert. IkeWiki - A Semantic Wiki for Collaborative Knowledge Management. Technical report, Salzburg Research, 2006.

[26] A. Souzis. Building a Semantic Wiki. *IEEE Intelligent Systems*, 20:87–91, 2005.

[27] S. Staab and R. Studer, editors. *Handbook on Ontologies*. International Handbooks on Information Systems. Springer Verlag, 2004.

[28] R. Tazzoli and P. C. et al. Towards a Semantic Wiki Wiki Web. In *Poster Session at the International Semantic Web Conference ISWC2004*, 2004.

[29] F. Vitali and M. Bieber. Hypermedia on the Web: What Will It Take? *ACM Computing Surveys*, 31(4):31, 1999.

[30] M. Völkel, M. Krötzsch, D. Vrandecic, H. Haller, and R. Studer. Semantic Wikipedia. In *Proceedings of the World Wide Web Conference WWW2006 (to appear)*, 2006.

[31] Wikimedia. Help:category. *http://meta.wikimedia.org/wiki/Category*.

[32] Wikimedia. Help:link. *http://meta.wikimedia.org/wiki/Link*.