# Semantic Wiki: Capturing Metadata within a Wiki Resource

**Dimitri Alexander Victor**
Department of Electrical and Computer Engineering
The University of Auckland
Private Bag 92019
Auckland, New Zealand
dvic005@ec.auckland.ac.nz

## ABSTRACT

This article describes the conceptual background of a Semantic Wiki, an augmented Wiki Wiki Web platform with technologies borrowed from the Semantic Web principles. The semantic wiki adapts a simple user interface to create wiki pages with semantic metadata as its underlying model.

The article will cover the basic anatomy of a semantic wiki, discussing its roots in the Semantic Web before venturing to analyse the core functionality and the architecture of the semantic wiki model. We will present some existing semantic wiki prototypes that are currently under development around the world. Finally we will discuss the challenges and limitations that must be overcome when dealing with the process of developing a semantic wiki platform.

### Author Keywords

Semantic wiki, wiki wiki web, semantic web, metadata, ontology, RDF, OWL, collaborative content editing.

### ACM Classification Keywords

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

## INTRODUCTION

A Wiki Wiki Web (or Wiki for short), is a website where users can contribute by adding content on any page. Cunningham [1] states that Wikis serve as repositories for ideas, providing an effective medium for discussion and collaborative work. The paradigms that used to exist between producers and consumers are becoming more blurred each day, helped by successful services such as Wikipedia, where anybody is able to contribute to a 'shared pool' of knowledge. Wiki's are becoming more and more popular in the business world as well, providing intranets and knowledge repositories in a simple and very affordable manner [2].

However the main limitation to the current Wiki Wiki Web model is that it simply consists of structured text and hyperlinks defining the data model. By adding a semantic model to the architecture, the semantic knowledge will be able to be processed by the system in a useful way, via the metadata attributes and relationships the users defined [3].

## BUILDING A SEMANTIC WIKI

### Towards the Semantic Web

The development of Semantic Wikis exists as part of the worldwide initiative towards creating a Semantic Web, deriving itself from W3C director Tim Berners-Lee's vision of the Web as a universal medium for data, information, and knowledge exchange [4].

The World Wide Web has become a vastly powerful tool for communication, research and commerce. However, its power is limited by the ability of human users to navigate various online resources to find the information they require. The Semantic Web vision is to make the Web machine-readable, allowing computers to integrate information from various sources to achieve the goals of end users. By augmenting web pages with descriptions of the content they hold, it becomes possible to reason about that content, resulting in advanced search techniques and generation of new facts from existing data. The potential impact of these technologies is huge, representing a reinvention of the world's computing infrastructure on at least the scale of the original web [5].

However distant the complete realisation of the Semantic Web vision might seem, the application of its capabilities and ideas can make software more connectable, interoperable and adaptable. Semantic technologies have the advantage of being designed with connectivity in mind, allowing different conceptual domains to work together as a network. Thus semantics-based schemas and tools have great potential to bring many benefits to existing systems and platforms [8].

### Semantic Wiki Functionality

Semantic Wikis take the principles of the Semantic Web and apply it to the existing Wiki model, where users can add semantic annotations to Wiki pages. These annotations are encoded in a formal language (RDF) such that the knowledge can be processed by the system. However the user does not necessarily need to know any particular RDF syntax to add semantic content, instead using the specified wiki syntax rules to edit semantic annotations [9].

Herman [5] states that the current Semantic Wiki developments have largely been focused on metadata and carefully designed data structures. The semantic knowledge stored in RDF can be machine-processed, allowing users to generate new facts and relationships about the stored data and query the wiki data model via the metadata.

To give an example, imagine a semantic wiki storing information on motion pictures. By storing semantic tags to such entities as films, actors, directors and relationships between them, we could effectively submit a query such as getting a list of "all thriller movies from the 1990's with New Zealand directors". So as you can see, there exists a great potential for formal reasoning based on the knowledge model obtained from a semantic wiki.

### Resource Description Framework (RDF)

A major component of a semantic wiki's model is the Resource Description Framework (RDF). RDF was designed to provide a common way to describe information such that it could be read and processed by computer applications.

RDF is a language that was designed for representing information about resources on the World Wide Web, particularly for metadata about Web resources. RDF is intended to provide a common framework for expressing information so it can be successfully processed by applications. RDF's ability to exchange knowledge across disparate applications means that the information may be available to applications other than those for which it was originally created [10].

The RDF model is based on the idea of making statements about resources in the form of subject-predicate-object expressions (called "triples"), establishing a particular relationship between the subject and the object. To understand how RDF triples function, consider representing the notion "The capital city of New Zealand is Wellington" as an RDF triple. In this case the *subject* is "New Zealand", the *predicate* is "capital city" and the *object* is "Wellington". The resources are described by URIs (Universal Resource Identifiers), and resource properties can either be resources or literals. (Eg. Wellington's population is 379,000)

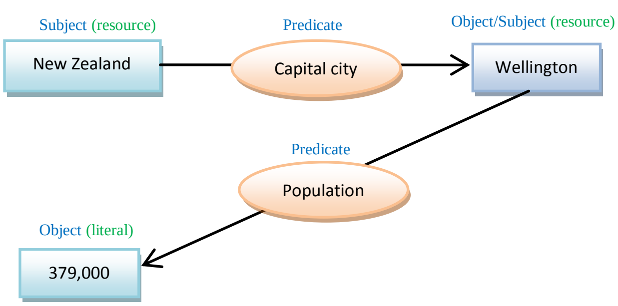


**Figure 1. A basic RDF model example.**

RDF documents are typically written in XML, using a syntax format called RDF/XML. Using a common standard ensures high levels of compatibility and interoperability, allowing computers to use different types of operating systems and application languages [5].

```
<?xml version="1.0"?>
<RDF>
<Description about="http://exampledb.com/newzealand">
 <capital resource="http://exampledb.com/wellington"/>
  ...
</Description>

<Description about="http://exampledb.com/wellington">
 <population="379,000"/>
 ...
</Description>
</RDF>
```

**Figure 2. The same model expressed in RDF/XML syntax. (Simplified example, namespaces are omitted)**

RDF additionally provides a platform for publishing both human-readable and machine-processable vocabularies. RDF's ability to provide a common standard for declaration of vocabularies encourages reuse of functions and creation of extensions to the RDF's data model among disparate information communities [10].

### Web Ontology Language (OWL)

Semantic wiki implementations also generally utilize an ontology language that provides an extension to the RDF syntax, supporting an enhanced vocabulary to illustrate the concepts of resources, properties, sub/super-classing, instantiation, inheritance and domain/range property restrictions [3].

The most common ontology language used for this task is the Web Ontology Language (OWL). OWL provides greater machine interpretability of Web content that is supported by XML, RDF and RDFS (RDF Schema) by providing additional vocabulary along with a formal semantics [11].

While semantic wiki technologies deviate from XML's ability to define customised tagging schemes and RDF's flexible approach to representing data, but at "ground level", semantic wiki's require an ontology language that can formally describe the vocabulary of semantic terms used in the system. OWL has been specifically designed to meet this need for a web-based semantic ontology language.

OWL consists of three sublanguages: OWL Lite, OWL DL and OWL Full, each used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms. OWL Lite supports users primarily needing a classification hierarchy and simple constraints. OWL DL is designed for users seeking maximum expressiveness while retaining computational completeness and decidability. While OWL Full is best fit for users requiring maximum expressiveness and syntactic freedom of RDF with no computational guarantees. Each of the sublanguages is an extension of its simpler predecessor, in terms of what can be legally expressed by the language syntax [12].

## EXISTING IMPLEMENTATIONS

There are several different implementations of semantic wikis that are currently under development around the world. More researchers are realising the immense potential that semantic web technologies have to offer to extend the functionality of the highly successful wiki model. The following are some of the existing semantic wiki prototypes in development including a brief description of the system.

### IkeWiki

IkeWiki is a semantic wiki prototype developed by SalzburgResearch, a non-profit research organisation of the State of Salzburg, Austria.

IkeWiki includes semantic annotations which can be added by users allowing some form of machine "understanding" of the information. The system allows context-specific presentation of pages and advanced querying functions [13].

Schaffert [14] writes that IkeWiki can also interpret some of the knowledge represented in its schemas to display enhanced navigational tools. For example, IkeWiki's interface automatically displays a taxonomy box for biological objects, if a biology ontology is accessed.

The purpose of IkeWiki is the transition from unstructured informal texts to a formal language taxonomy. The tool aims to provide domain experts who are not familiar with complicated tools and languages (such as Protégé and OWL), to be able to formalise their domain knowledge to be accessible by other Semantic Web applications [14].

### Semantic MediaWiki

Another prominent semantic wiki prototype is the Semantic MediaWiki, a semantic extension to the popular MediaWiki platform, famous for being the wiki platform for the highly acclaimed Wikipedia.

Völkel et al [6] describe Semantic MediaWiki as an extension that enables wiki-users to semantically annotate wiki pages, which then allows for the content to be browsed, searched and reused in new innovative ways. The overall objective of the project is the integration of the Semantic Web technologies into Wikipedia through the development of a solution for semantic annotation that fits the needs of most Wikipedia projects and still meets the requirements for usability and performance.

With the semantic model to structure the data, Wikipedia has the potential to include more specific ways of searching and browsing its information. And also this formalised knowledge, would allow external applications to automatically access Wikipedia's knowledge base, for its own specific needs [6].

'''London''' is the capital city of [[England]] and of the [[United Kingdom]]. As of [[2005]], the total resident population of London was estimated 7,421,328. Greater London covers an area of 609 square miles. [[Category:City]]

'''London''' is the capital city of [[capital of::England]] and of the [[is capital of::United Kingdom]]. As of [[2005]], the total resident population of London was estimated [[population:=7,421,328]]. Greater London covers an area of [[area:=609 square miles]]. [[Category:City]]

**Figure 3. Source of a Wikipedia page on London using current markup (top) and with semantic extensions (bottom). (Adapted from [6]).**

### Kaukolu Wiki

Kaukolu Wiki is a semantic wiki research prototype used as a testbed for alternative approaches of mapping wiki contents to a knowledge base. Kaukolu is based on JSPWiki, a wiki engine build using the standard J2EE components (Java, JSP, servlets) [15].

While most semantic wikis work by identifying content pages with respective RDF resources and links between pages with RDF predicates, Kaukolu implements an indirection layer between the wiki page model and the RDF model. RDF resources can get created and may be 'attached' to sections of text, forming annotations. Users can navigate both in the standard wiki content and the layer formed by annotations, switching between both views as needed [16].

Another interesting thing about Kaukolu is that the users can create semantic annotations using a form-based editor featuring a GUI interface, whereas most other semantic wiki systems use a more conventional text-based input.

### OntoWiki

OntoWiki uses a wiki concept to implement an ontology development environment, providing users with simple editing tools for creation and maintenance of ontologies.

OntoWiki uses multimedia elements to improve the richness of the tool, facilitating the visual presentation of a knowledge base as an information map, with different views on instance data. The tool enables users to create and edit semantic content via an intuitive GUI interface, making the system relatively simple to use.

OntoWiki aims to promote aspects of social collaboration between users, by featuring functionality for discussion space and user comments in every part of the knowledge base. The tool also provides users with semantic search capabilities, improving searching and retrieval of information stored in the system [17].

## APPLICATIONS OF THE TECHNOLOGY

The single-most important application of the semantic wiki technology is the *integration* of knowledge, such that the semantic data could be processed by systems around the world to enhance the functionality and provide an overall benefit to the application. This concept of integration includes seamless incorporation of the semantic wiki data into desktop applications, web content, business databases and other various computer systems [3].

There are many desktop applications which could be enhanced by integrating data from a semantic wiki's knowledge base. One of the examples of such integration would be accessing content for desktop media playback applications, greatly improving the current capabilities of media tools. For instance, with semantic wiki integrated into the media suite, we could retrieve a great amount of information regarding music artists, records, films and other media, which could be processed to provide up to date information on the selected media, as well as vastly improve the search capabilities for related media items [6]. These integrated media tools could answer domain-specific queries such as "music artists influenced by Crowded House" or "films that received an Academy Award and have Peter Jackson as the director". Recommender systems for these media tools could also become far more advanced as a consequence of the semantic knowledge.

The semantic integration of desktop applications is not restricted to the realm of media tools. Most of today's applications could benefit from data acquired from a semantic wiki knowledge base. Word processing applications could be enhanced to gather information on any topic as well as recommending related material that could be relevant. Image processing applications could incorporate tools to search for images on certain topics. Spreadsheet applications could automatically visualise the data retrieved from a semantic wiki. The opportunities for useful amalgamation of the conventional desktop tools with semantic content are endless.

The application of semantic wiki technologies is also very relevant to Web content. A web-based interface could as easily apply the same principles that were discussed for any of the above systems. Search engines could utilize the acquired semantic data to provide search techniques that are more efficient and robust. The use of semantic content would also provide internationality of data, allowing the semantic machine-processable data to be used by communities across the globe, with the ability to translate the semantic knowledge into different languages, using existing open-source wiki tools (eg. Wiktionary) [6].

The future offers endless possibilities for integration of semantic wiki data into a multitude of diverse computer applications that will greatly benefit the end-users of these systems.

## IMPLEMENTATION CHALLENGES

There are a number of obstacles standing in the way of implementing an effective semantic wiki system. We will discuss the most relevant challenges that must be overcome in order to create a successful semantic wiki platform.

### Usability

The first challenge is satisfying the highest requirements on usability, as the wiki model relies on a large community of volunteers for authoring/editing the content. By adding extra syntactic rules to establish semantic content of the data, we can unwittingly start to counteract the ease of use of wiki editors, hence losing one of the key advantages of using a wiki in the first place [9].

Thus it is absolutely imperative for users to be able to use the extended semantic features without any specific technical knowledge or training. Another idea would be to have the semantic content editing as an optional feature in the knowledge base. This strategy would result in maintaining any existing wiki community and will also allowing users to add semantics to the existing content.

### Scalability

Oren et al [7] state that a major obstacle is the sheer large scale of the systems. The number of link-types that must be present in a comprehensive semantic wiki in order to create a good ontology is quite large, hence creating the problem of choosing the correct link type for each relationship and so on. A successful semantic wiki implementation must be able to provide exceptional performance and be able to scale well, as its community and knowledge base expand as the system grows.

There will also be some problems in the realm of scalability resulting from making changes to the knowledge base that involve a number of different subjects. This will mean that all the corresponding wiki content will need to be parsed and updated accordingly [8]. To be truly functional any successful semantic wiki implementation must in-part provide some level of automation when it comes to the content being updated. Semantic knowledge in the system must at all times match the human interface content that is available.

### Flexibility

The wiki platform is extremely flexible, being able to support a diverse variety of tasks, providing users with almost no restrictions regarding the type of content and its presentation. Semantic wiki's will need to have a much stricter structure and this may impose some restrictions on the editing community. A successful semantic wiki implementation must allow the users a good sense of flexibility in terms of content editing as well as providing a formal semantic vocabulary model.

### Expressiveness

How much knowledge should we have in a machine-processable form? It would be desirable to have as much metadata content as possible in a semantic wiki, however as we increase the amount of semantic content we will start to experience conflicts with the platform's usability and performance. Difficulties such as logical inconsistencies and previously mentioned problems with large number of link types need to be avoided [18].

### Compatibility

Ensuring compatibility with current tools and technologies is crucial to a successful semantic wiki implementation. This includes compatibility with the current tools and services as well as any future developments that may arise. By introducing any level of incompatibility we may compromise the community support behind the system [9]. To ensure this does not happen, there must be careful selection of appropriate semantic description languages and tools that meet the users' compatibility standards.

## CONCLUSIONS

Wikis have definitely already proved themselves to be an excellent storage medium for knowledge management, providing users with fast, easy access to a vast set of knowledge. By including semantic metadata to the wiki format it can be possible to perform formal reasoning on the stored information, allowing the users to get much more usability from the data stored in the system.

Semantic wikis derive themselves from existing Semantic Web technologies, in particular RDF and OWL. Currently there are numerous semantic wiki implementations that are striving towards the goal of developing a successful semantic wiki platform. Semantic wiki designers are faced with numerous design challenges that need to be overcome. However, the emerging pool of machine-accessible data presents greater opportunities each day for developers to overcome the challenges and employ these semantic tools in practice.

## REFERENCES


1. Cunningham, W., Design principles of wiki: how can so little do so much?, *Proceedings of the 2006 International Symposium on Wikis,* (2006), 13-16.
2. Krotzch et al, Wikipedia and the Semantic Web The Missing Links, *Proceedings of Wikimania*, (2005), Frankfurt, Germany.
3. Tolksdorf, R., Simperl, E., Towards Wikis as Semantic Hypermedia, *International Symposium on Wikis*, *ACM* (2006)
4. Berners-Lee, T., Hendler, J., and Lassila, O., The Semantic Web. *SA*, *W3C* (2001)
5. Herman, I., W3C Semantic Web Activity, *W3C* (2008).
6. Völkel et al, Semantic Wikipedia, *WWW '06: Proceedings of the 15th international conference on World Wide Web, ACM* (2006)
7. Oren et al, How semantics make better wikis, *WWW '06: Proceedings of the 15th international conference on World Wide Web, ACM* (2006)
8. Maedche, A., Staab, S., Ontology Learning for the Semantic Web, *IEEE Intelligent Systmems*, (2001), 72-79.
9. Souzis, A., Building a Semantic Wiki, *IEEE Intelligent Systems*, (2005), Vol.20, No. 5, 87-91.
10. RDF Primer. W3C Recommendation 10 February 2004. http://www.w3.org/TR/REC-rdf-syntax/.
11. Horrocks et al, From SHIQ and RDF to OWL: the making of a Web Ontology Language, *Web Semantics: Science, Services and Agents on the World Wide Web*, (2003), Vol.1, Issue.1, 7-26.
12. OWL Web Ontology Language. W3C Recommendation 10 February 2004. http://www.w3.org/TR/owl-ref/.
13. IkeWiki. Info Page. http://ikewiki.salzburgresearch.at/
14. Schaffert, S., IkeWiki: A Semantic Wiki for Collaborative Knowledge Management. *Enabling Technologies: Infrastructure for Collaborative Enterprises, WETICE '06*, (2006), 388-396.
15. KaukoluWiki. About. http://kaukoluwiki.opendfki.de/
16. Kiesel, M., Kaukolu: Hub of the Semantic Corporate Intranet. DFKI GmbH, Kaiserslautern, Germany, (2006)
17. Auer et al, OntoWiki – A Tool for Social, Semantic Collaboration. *International Semantic Web Conference,* (2006)
18. Aumueller, D., Semantic authoring and retrieval within a wiki. *Proceedings of the European Semantic Web Conference*, (2005)