



MARCUS SIMON

TEACHING THE OLD WEB NEW TRICKS

Ever since the advent of the online encyclopedia Wikipedia, the average Web user knows what a wiki is. The term 'semantic wiki', on the other hand, is likely to evoke a quizzical look from most people. It refers to a tool that combines the wiki idea with the advantages of the Semantic Web. To find out what the concept entails and to understand its potential applications, one only needs to take a look at the work of LMU computer scientists Professor François Bry and Klara Weiland.

"Pretty." – "Yeah, just look at those dead bastards." This cynical exchange between US troops can be heard on a video that shows how a combat helicopter in Iraq fires on and kills a group of civilians, who posed no obvious threat to the military. Among the victims were two reporters for Reuters news agency, whose cameras may have been mistaken for weapons by the soldiers. This video turned up on the Internet in April 2010, and was highly embarrassing for the US Army. Any users who had not come across the word 'wiki' before will have encountered the term on this occasion. For the video appeared on Wikileaks, an online platform that gives insiders with access to sensitive information the opportunity to make it available worldwide, instantaneously and – anonymously. A better known, and much less controversial, wiki platform is the online encyclopedia Wikipedia. Some 350 million Internet users make use of Wikipedia every month. Since the German language version went online in 2005, a growing number of users in this country have learned that a wiki is a system that enables users to write and edit content on the Web. This type of communal work on evolving content is now referred to as "collaborative production of knowledge." But if the idea of a wiki is no longer a mystery to Web users, what exactly is a semantic wiki? This extension of the wiki concept will not yet mean much to most Web users. This is strikingly confirmed if one consults the Web's own resources. Google lists about 40,000 items that contain the term – not very impressive in comparison with the "approximately 353 million results" one gets if one uses the unadorned 'wiki' as a search term. A small coterie of informatics experts is working to change this situation. Professor François Bry is one of them. "Wikis need moderators to coordinate the work", says the Head of the Research

Unit for Programming and Modelling Languages in the Department of Computer Science at LMU, “but they provide a relatively easy means to incorporate improvements into existing documents.” If an “improvement” turns out to be nonsensical or useless, the moderators can intervene. A wiki works on the principle of ‘first trust and then check’. As a tool, it is particularly suitable for bringing together the work of widely distributed teams of experts belonging to firms that operate worldwide. A group of architects, for instance, might use a wiki to exchange information about legal regulations, building regulations and prices, say, in a city in which they wish to realize a planned project. Individual users can be provided with different rights of access. Some might be allowed to view, but not to alter, data regarding financing, others might be denied access to those pages altogether. “The idea of the semantic wiki is simply that one adds a modicum of – built-in – intelligence”, says François Bry.

MACHINES MUST LEARN TO “UNDERSTAND”

“The Semantic Wiki brings together two developments that have occurred over the past few years”, remarks Klara Weiland, a doctoral student in Bry’s group. “The first is the evolution of the Social Web, where most users are not just passive consumers of content, but also supply content.” Wikis are one example; blogs and platforms like Facebook also fall into this category. “The second development is the Semantic Web.” In order to grasp what this involves, it is helpful to consider the World Wide Web as we now know it. We can use search engines to find web pages on the basis of keywords. But the information we find by this means cannot be logically linked together automatically, it is a simple list of web sites. – We cannot use such a procedure to obtain answers to questions, because the semantic content of the pages it finds is not accessible to the search engine. Google, Bing et al. can only answer one type of question: “Which web pages contain the term X?”. Maybe the question we really wanted to have answered when we initiated the search was: “Where can I find, within a radius of 3 km of Geschwister-Scholl-Platz, a shop where USB sticks cost less than 10 Euros?” We can interpret symbols that encode information, and connect them contextually with a specific semantic content – our machines do not possess this ability. Machines can transmit and manipulate symbols received from other machines in a given system, but they cannot interpret them in terms of their meaning in the world beyond the system. – This is where the Semantic Web comes in. The basic idea is that it should be possible to assign a meaning to content based on semantic context. The software should be able to recognize whether, for example, ‘COAT’ is being used as a noun or a verb, and whether ‘COOK’ refers to a surname, an action or a job title. In other words, the Semantic Web is an attempt to “teach” machines to “understand” facts. The facts must be expressed so unambiguously that no further user input is required for the machine to interpret them. This would drastically reduce the amount of irrelevant information that search engines turn up, and allow software “agents” to perform tasks entirely autonomously. In order to realize this vision, one need not replace the existing architecture of the Web, one adds an extra level to it. This level must operate on metadata – information about information – and be capable of

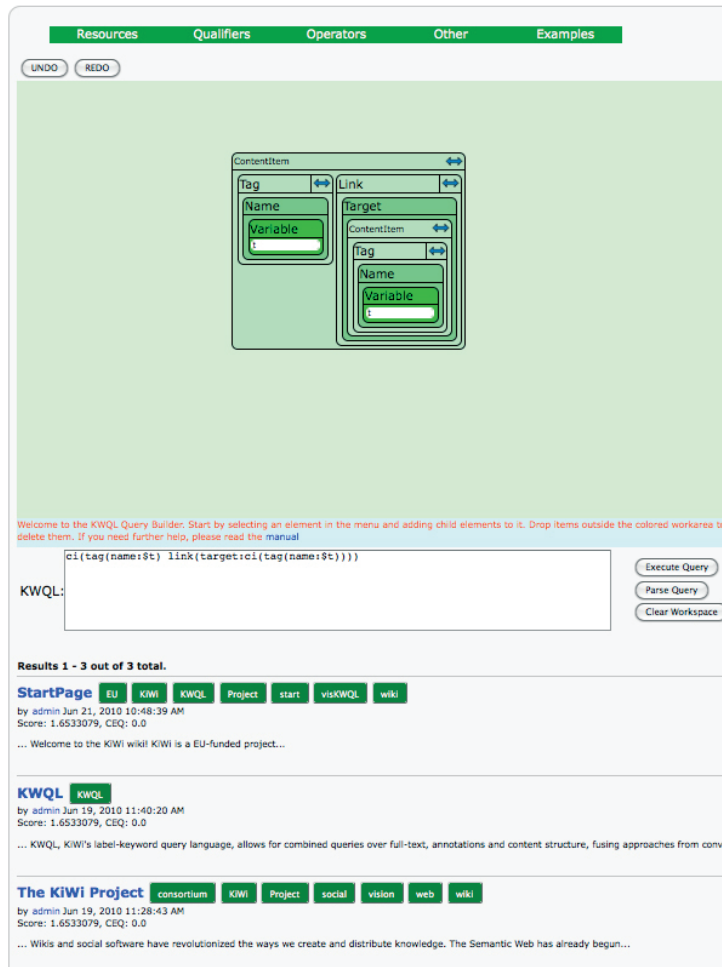
```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <foaf:Person>
    <foaf:name>Monika Mustermann</foaf:name>
    <foaf:mbox rdf:resource="mailto:monika.mustermann@deutschland.de" />
    <foaf:homepage rdf:resource="http://www.monika.mustermann.de" />
    <foaf:knows>
      <foaf:Person>
        <foaf:mbox rdf:resource="mailto:myfriend@example.com" />
        <rdfs:seeAlso
          rdf:resource="http://example.com/~myfriend/foaf.rdf" />
        </foaf:Person>
      </foaf:knows>
    </foaf:Person>
  </rdf:RDF>

```

This example shows an individual's description in FOAF, including the term 'Element knows', which links up to another person's FOAF file.
Source: LMU Munich

structuring data in meaningful ways. In the context of wikis, this implies that content is first collected and then structured. The structuring step can be hierarchically organized, but the perspective on a particular item or case should not be rigorously defined or otherwise restricted in advance. "One of the greatest unsolved questions in the field is how queries can refer to the documents' structure", says François Bry. "The central problem is the size of the Internet. If I were to ask a search engine like Google, which operates on web sites and their associated links, to find a site that contains the term A and is linked in a specifically defined sense B to a site that contains information on C, Google would be stymied. It simply cannot process such a query." However, using appropriate algorithms, it is possible in a semantic wiki to combine queries with semantic structures and a ranked list of possible answers. "The essential idea is that one employs formalisms to express the content of texts in a form that the computer can interpret", explains Klara Weiand. "So we have the wiki aspect plus annotations that make the semantic connotations of a text "legible" for the computer, so that the machine can automatically distill facts (relationships with other texts) from a given content." The content is supplied by enthusiastic users. The Semantic Web itself is focused on the representation of knowledge, innovations and on distributed collaborations involving project groups with or without a hierarchical organization. As François Bry emphasizes: "If following 'prescribed procedures' is the only way to get things done, then an organization reacts sluggishly and is inefficient." Ideally, knowledge management should be able to make use of implicit knowledge of processes and distributed knowledge archives, and combine semantically diverse types of content, such as source code, documentation, exercises and plans in useful ways. The systems now available do not adequately fulfill these criteria. They are not flexible enough to integrate all of these areas and make them accessible in a useful form. To remedy this situation, the European Union has provided 2.69 million Euros to fund a 3-year project entitled "Knowledge in a Wiki" (KiWi). The participating research-



The query language visKWQL allows one to enter not only simple textual queries (lower field) but also visual queries (upper field, green background)
 Source: LMU Munich

be extracted automatically from text. A text is analyzed and a user can choose from a set of annotations the one that best describes the content for the intended purpose. The researchers at Aalborg University are looking at personalization – recommending to the current user web sites that might be of particular interest. This feature allows the wiki to be tailored to the needs of individual users.” The team at LMU Munich is concentrating on the formalization of queries and reasoning – enabling the computer to make inferences from content. Klara Weiland’s project tackles the problem of how one might intelligently connect simple keyword queries with the syntactic and semantic structure of content. She is developing a language for the formulation of queries within a semantic wiki, called KWQL (pronounced “quicquel”). “Conventional web query languages can be used for querying data”, she says, “but that in itself is actually quite complicated, as one needs to know both the data to be searched and the query language itself pretty well.” On the other hand, web search allows the user to simply type in a keyword or a set of keywords that occur together in the type of content one is looking for. But, as Klara Weiland points out, “keyword queries can only roughly approximate the user’s query intent. The basic goal of my work is to bring web querying and web search together, to enable users to target searches much more precisely. In a semantic wiki one has text, metadata and data structures. This means that a particular

ers plan to design a working “KiWi system” for knowledge management, based on the use of a semantic wiki. The system is intended to support collaboratively assembled knowledge archives and will combine these with Semantic Web technologies. The KIWI project is being coordinated by Salzburg Research, an Austrian research agency that has already developed “IkeWiki”, an open-source software package for semantic wikis. In addition to LMU Munich, Aalborg University (Denmark) and Brno University of Technology (Czech Republic) are partners in the project. “The team in Brno is working on information extraction”, says Klara Weiland. “Its task is to design ways in which proposals for annotations can

section of a wiki page is located on a given level, or that certain sites link to each other. So one has various types of data to consider. The goal is to find ways that allow normal users to profit from all types of data structures without having to become experts in query languages first." She hopes that her work will make it possible to pose a sequence of increasingly complex, expressive and specific queries, starting from a simple keyword search, so that the end result comes pretty close to answering the question one had in mind at the outset. "The KiWi system is very user oriented, so one cannot make the task of formulating queries too demanding." If you would like to try the system out, please do so at: <http://www.pms.ifi.lmu.de/kwql/> There you will find a system that can handle visual as well as text-based KWQL queries (see figure on p. 4).

As Klara Weiland explains, there are still many hurdles to be overcome. "The aim is to make it as easy as possible for every user to enter data in semantic wikis. Very powerful, but very complicated, formalisms for generating semantic annotations already exist. The problem is that one needs experts to write them down. If a formalism is too complex, very few users will be willing to contribute content." Since the goal is to construct a platform for social knowledge management based on a semantic wiki, it is of paramount importance that the system be user friendly. This is why the concept of a user-centred system is at the heart of the KIWI project. "RDF is a language for knowledge representation which employs a formalism that is relatively easy to use", says François Bry. RDF stands for Resource Description Framework and offers a way of describing Web objects and their relationships to one another. "RDF operates with the expressions Subject, Predicate and Object. In our wiki, every web site is a Subject, and Predicate and Object are attached as annotations." The application Friend of a Friend (FOAF) makes use of RDF. FOAF describes people, their relationships and their activities. Several tools are available which permit one to produce FOAF files that conform to FOAF specifications using graphic interfaces. The FOAF file is then made accessible to Web users by means of a standard symbol string, the Uniform Resource Identifier (URI). The URI is a description of a real person. Simply by entering the URIs for different individuals, one can automatically ascertain whether, and via which mutual acquaintances, the individuals are linked to each other (see the figure on p.3).

François Bry can also list other applications. "NATO uses wikis at command level. At the other end of the chain of command, wikis are being used by the troops in Iraq." Ground patrols use them for rapid exchange and transmission of information. Relaying this kind of information through "official channels" would take too long, and one could not warn other units of unexpected obstacles or uneasy situations in a particular location. Modes of semantic enrichment for the wikis used by US troops are being actively sought. Wikis are also finding more and more applications in commercial firms. Semantic wikis are very popular in firms that develop and market software, especially those that tap into and utilize open-source programs. Other high-tech industries, such as the aerospace industry, in which cycles of innovation succeed each other very quickly, also use wikis. Why is that

so? François Bry has a clear and simple answer to that question: these are firms that must remain competitive in markets that are changing at a dizzying pace.

Translated by Paul Hardy

Prof. Dr. François Bry has held the Chair of Computer Science at the Department of Computer Science since 1994. Together with Prof. Dr. Hans Jürgen Ohlbach, he directs the Research Unit for Programming and Modelling Languages at LMU Munich.

bry@lmu.de

<http://pms.ifi.lmu.de/>



Klara Weiand joined the Doctoral Program at the Research Unit for Programming and modelling Languages in 2008. Her thesis work is part of the EU-funded project "Knowledge in a Wiki".

http://www.pms.ifi.lmu.de/mitarbeiter/klara-weiand?set_language=de

klara.weiand@ifi.lmu.de

