Semantic Wikis Versioning with BiFröST

Krzysztof Kutt[*]

AGH University of Science and Technology
Al. Mickiewicza 30, 30-059 Krakow, Poland,
kkutt@agh.edu.pl

Abstract. One of the powerful and popular tools that are used to support Collaborative Knowledge Engineering are Semantic Wikis. They are easily accessible and provide ACL mechanisms, but they lack a good versioning mechanism. In this paper an extended version of such a mechanism is proposed. Besides elements that appear in every Wiki system, like simple changelog and place for discussion, it incorporates changes ontologies, rich metadata, semantic metrics and reasoning unit tests. All of them are gathered into the form of the provenance graph that can be serialized into Turtle syntax and (automatically) analyzed. First prototype of such a mechanism for Loki wiki called BiFröST was developed.

Keywords: semantic wikis, collaborative knowledge engineering, knowledge evaluation, change management

1   Introduction and Motivation

Knowledge bases are created not by a single knowledge engineers, but by a team of people who either sit in the same room (in a company) or people who do not even know each other (as in the Wikipedia project[1]). It can therefore be easily seen that there is a need to move from traditional Knowledge Engineering (KE) to Collaborative Knowledge Engineering (CKE) and to consider issues related to the collaborativeness.

Primarily, following issues should be taken into account, cf. [12]: (1) The sources vary in the knowledge of the topic, in the same way the user do. Is it possible to identify which of them are better? (2) Conflicts in collaborative setting are natural. People in general and experts in particular may have different views on the same subject. Conflicts can also arise from the bad will of some users: they can consciously enter incorrect information to the knowledge base, e.g. vandalism on Wikipedia. How can we resolve conflicts? (3) There are different kinds of users and different types of changes: some users aim at enhancing the knowledge base, the other ones at bringing coherence, and still others at improving text quality. Is there a way to identify them and use in some way?

Among tools that support CKE teams and tries to address these issues are: (a) WebProtégé[2], a collaborative version of open source ontology editing tool

---

[*] The author pursues his doctorate under the supervision of Grzegorz J. Nalepa.
[1] See: https://www.wikipedia.org/.
[2] See: http://webprotege.stanford.edu/.

Protégé that was used, inter alia, in the process of the International Classification of Diseases (ICD-11) preparation, where 270 doctors have worked together on the base that consists of 45k classes [14], (b) Noctua[3], another online tool, which provides knowledge validation by interacting with users. System creates user profiles and then asks questions related to each one's field of knowledge or interest (e.g. "Could you write something about X?") [2], and many more.

In this paper, one specific group of tools that supports CKE teams will be considered. They are called Semantic Wikis and they are quite interesting because of their popularity (see section 2). Like traditional wikis, they are conceptually simple but powerful for massive cooperation, and also provide a semantic layer on which one can define concepts and relations that are easily machine-understand.

What is important from the CKE point of view, is the fact that every Wiki system has integrated collaboration mechanisms such as ACL (Access Control Lists), which allows for gaining different access levels for different users and version control for tracking all changes that was done within wiki pages. Among Semantic Wikis' drawbacksis the fact that they provide only simple "straight" changelogs in which all versions are saved with small comments, as well as some place to discuss editions, but nothing more. This is not enough to really support CKE within wikis, because "classical" changelog does not provides any way to specify used resources nor to automatically evaluate conflicts.

In this paper extended versioning mechanism for Semantic Wikis is proposed with taking care of CKE collaborative issues presented above. It is based on a semantic graph in which changes are related to each other as well as to some additional information, what gives a possibility to automatically analyse the changelog and provide some feedback about collaboration quality.

The rest of the text is structured as follows. In part 2 Semantic Wiki systems and their capabilities are characterized. Section 3 outlines the approach. The paper is concluded in section 4.

2   Semantic Wikis

Wiki systems are a kind of containers for text pages ("wiki pages") and media files (images, audio files, and others). Wiki pages are human readable plain text files that uses special structures to denote different types of content (e.g. headers, images). Pages within a Wiki are usually grouped within the so-called namespaces and identified by unique names. They are also linked to each other creating a hyperwikitext structure[4].

Semantic wikis enrich the wiki technology with semantic information. They are quite popular, as evidenced by the large number of semantic wiki systems. Examples are: KnowWE [1], OntoWiki [6] or SweetWiki [3]. The most popular is the Semantic MediaWiki (SMW) [7] that is based on the MediaWiki system used by the Wikipedia project. Its capabilities include, but are not limited to, stating information about categories and relations between wiki pages and querying a

---

[3] See: `http://projetos.dia.tecpar.br/noctua/index.php?clicked=info`.

[4] For more information and wikis comparison see: `http://www.wikimatrix.org/`.

wiki with a specific query language as well as with a standard SPARQL language. Semantic Wikis are therefore good knowledge management tools [8] that are used not only in KE projects but also in software engineering ones [4,5], [10,11].

Loki [8,9] is another Semantic Wiki, developed as a plugin for DokuWiki system that stores knowledge using Prolog-based representation. It is distributed as free software[5], what makes it easily accessible by any interested group of people. To be more precise, Loki is not a single plugin but a set of independent plugins, e.g. BPwiki [10] for storing and visualizing Business Processes within wiki and SBVRwiki [11] for processing business rules.

Semantic Wikis were used in many projects, ranging from historical ones, like Catalogus Professorum Lipsiensis, knowledge base that describes life and work of 1.400 professors of Universität Leipzig in 1409-2009 [13], through the business ones, e.g. Prosecco project, where semantic wiki was used as a storage for SBVR business rules that describes the small and medium enterprises [11], to medical ones, like Digitalys CareMate, a decision support system for paramedics [1].

3   Outline of the Solution

Extended versioning mechanism for Semantic Wikis proposed in this paper is based on the PROV graph[6], a W3C standard prepared for describing data provenance. PROV graph represents relations between entities (wiki pages and their revisions), activities and agents (users). E.g. the graph states that the `first revision` of a page `welcome` was created on `31/08/2016` by `Chris` (what consists of three triples that connect selected wiki page–subject, with main page, creation time and author)[7]. New facts (triples in the RDF form) are added to the changelog after one of three possible user activities: page creation, edition and deletion. The graph incorporates also pieces of information provided by modules, which were selected in order to most comprehensively catch the properties of the changes (see Figure 1):

Changes ontologies: main activities within Wikis are page creation, edition and deletion, but these ones do not exhaust the issue. We should take into account that there are different types of changes (see Section 1) and each edition is stated by the factual change (**What** was done? E.g. *Typos or other small bugs fixed* or *New content added*) and the goal (**Why** it was done? E.g. *Errors fixing* or *Knowledge database expansion*) that form some kind of ontologies (one for every task's "type", e.g. one for preparing conference papers and the other for developing system specification within wiki).

Rich metadata: besides "standard" timestamp and user, the framework also provide more information about each page edition: internal (other wiki pages) and external (e.g. books) sources of knowledge in the form of an URI, as well as about user: name, e-mail address, etc.

---

[5] See: `http://loki.ia.agh.edu.pl/`.

[6] See W3C documentation on: `https://www.w3.org/TR/prov-primer/`.

[7] Detailed specification of created triples is provided on the page: `http://loki.ia.agh.edu.pl/wiki/docs:prov#how_the_prov_plugin_works`.
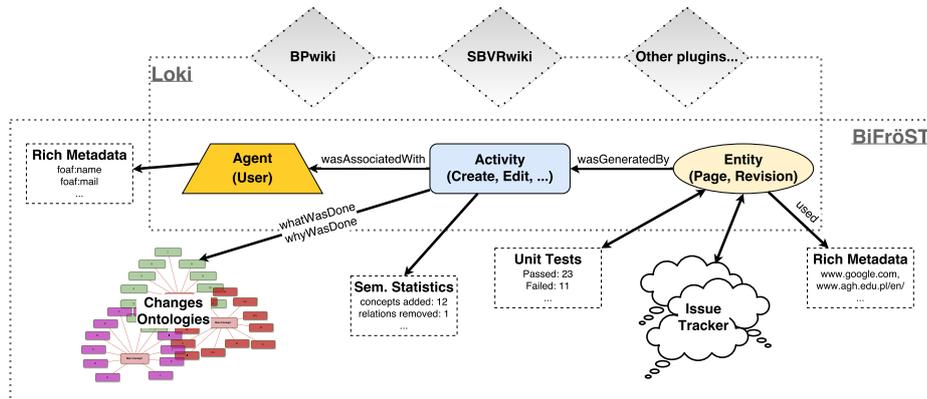
Fig. 1. Extended semantic wiki versioning architecture. The BiFröST framework is a prototypical implementation based on the Loki Semantic Wiki.

Semantic statistics: how many concepts / instances / relations were added / removed / corrected in this revision?

Reasoning (unit) tests: as an analogy to the unit tests in software engineering, we added a possibility to define test cases that will be executed during page saving to ensure the minimum level of knowledge quality.

Issue tracker: it is not possible to cover all aspects of changes by automatically computed characteristics (semantic statistics and reasoning tests mentioned above). There should be also a place for discussion between experts.

The advantage of the proposed solution over "classical" changelog within CKE setting is based on the power of the PROV graph based on the RDF triples that can be (automatically) queried with SPARQL. With this in mind exemplary use case scenarios are indicated, as a response for issues described in Section 1:

Get rid of bad changes. With tests' statistics, one can quickly identify which changes are bad and should be examined. In a more rigorous case, a wiki's administrator can block the ability to save changes if new revision is worse than previous one (less tests were passed).

Sources analysis. By combining tests' statistics and sources lists, to determine which sources have low quality and shouldn't be used in the future.

User types identification. Thanks to placing the solution on the top of changes ontology, we can identify different kinds of users, e.g. good users and bad users (who introduces bad changes) or creators (they add a lot of text), annotators (they provide many new relations for existing text) and proofreaders (users that mainly fix spelling).

Underdeveloped pages indication. Semantic statistics provide information about how many concepts and relations are stated on the page. If there are not too much, maybe it is a good time to pay attention to the page?

Motivation by gamification. Accurate metrics allow for awarding points to users, giving them badges for different tasks (e.g. badge for adding 5 new

relations in the wiki pages) and creating leaderboards. These ones can motivate them to create better knowledge base what is the main goal.

The BiFröST (**Bi**FröST **Fr**amework f**ö**r **S**emantical **T**racking), a prototypical implementation of proposed versioning mechanism for Loki Semantic Wiki is being developed and tested in collaboration with students at AGH-UST[8]. At the time of submitting the paper, PROV graph is created within framework (see Figure 2). This first prototype was used in small experiment to collect triples about 940 user activities. Practical implementation of the SPARQL queries scenarios presented above is a future work.
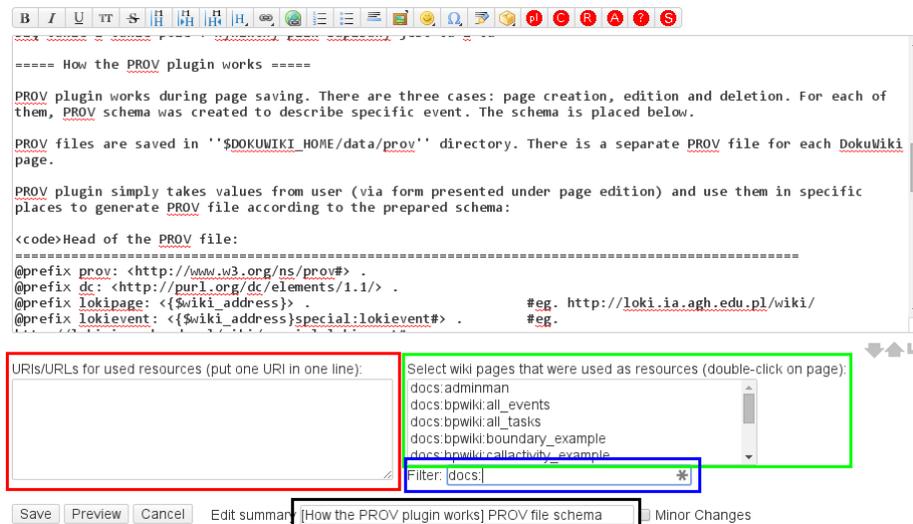
Fig. 2. First BiFröST prototype. Figure presents form that consists of: (a) place for URIs/URLs for used resources, both internal and external (red frame); (b) all wiki pages index (green frame) with filter capabilities (blue frame) for easy internal resources insertion; (c) whatWasDone and whyWasDone fields for selecting proper concepts from changes ontology (not visible on the figure).

4   Summary

In this paper the enhanced semantic versioning mechanism is proposed as well as prototypical implementation called BiFröST. Besides elements that appear in every Wiki system, like simple changelog and place for discussion, the proposal incorporates changes ontologies, rich metadata, semantic statistics and reasoning unit tests. All of them are gathered into the form of the provenance graph that can be analyzed using SPARQL queries. First prototype implementation was developed and used in small experiment at AGH UST.

---

[8] See: `http://loki.ia.agh.edu.pl/wiki/docs:prov`.

References

1. Baumeister, J., Reutelshoefer, J., Puppe, F.: Knowwe: A semantic wiki for knowledge engineering. Applied Intelligence pp. 1–22 (2011), `http://dx.doi.org/10.1007/s10489-010-0224-5`, 10.1007/s10489-010-0224-5
2. Boz, G., Ramos, M.P., Sato, G.Y., Nievola, J., Paraiso, E.C.: Noctua: A tool for knowledge acquisition and collaborative knowledge construction with a virtual catalyst. In: Computer Supported Cooperative Work in Design (CSCWD), 2011 15th International Conference on. pp. 222–229. IEEE (2011)
3. Buffa, M., Gandon, F.L., Erétéo, G., Sander, P., Faron, C.: Sweetwiki: A semantic wiki. J. Web Sem. 6(1), 84–97 (2008)
4. Decker, B., Ras, E., Rech, J., Jaubert, P., Rieth, M.: Wiki-based stakeholder participation in requirements engineering. Software, IEEE 24(2), 28–35 (March 2007)
5. Dengler, F., Vrandečič, D., Simperl, E.: Comparison of wiki-based process modeling systems. In: Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies. pp. 30:1–30:4. i-KNOW '11, ACM, New York, NY, USA (2011)
6. Frischmuth, P., Martin, M., Tramp, S., Riechert, T., Auer, S.: Ontowiki–an authoring, publication and visualization interface for the data web. Semantic Web 6(3), 215–240 (2015)
7. Krötzsch, M., Vrandecic, D., Völkel, M., Haller, H., Studer, R.: Semantic wikipedia. Web Semantics 5, 251–261 (2007)
8. Nalepa, G.J.: Collective knowledge engineering with semantic wikis. Journal of Universal Computer Science 16(7), 1006–1023 (2010), `http://www.jucs.org/jucs_16_7/collective_knowledge_engineering_with`
9. Nalepa, G.J.: Loki – semantic wiki with logical knowledge representation. In: Nguyen, N.T. (ed.) Transactions on Computational Collective Intelligence III, Lecture Notes in Computer Science, vol. 6560, pp. 96–114. Springer (2011), `http://www.springerlink.com/content/y91w134g03344376/`
10. Nalepa, G.J., Kluza, K., Ciaputa, U.: Proposal of automation of the collaborative modeling and evaluation of business processes using a semantic wiki. In: Proceedings of the 17th IEEE International Conference on Emerging Technologies and Factory Automation ETFA 2012, Kraków, Poland, 28 September 2012 (2012)
11. Nalepa, G.J., Kluza, K., Kaczor, K.: Sbvrwiki a web-based tool for authoring of business rules. In: Rutkowski, L., [et al.] (eds.) Artificial Intelligence and Soft Computing: 14th International Conference, ICAISC 2015: Zakopane, Poland. pp. 703–713. Lecture Notes in Artificial Intelligence, Springer (2015)
12. Richards, D.: Collaborative knowledge engineering: socialising expert systems. In: 2007 11th International Conference on Computer Supported Cooperative Work in Design. pp. 635–640. IEEE (2007)
13. Riechert, T., Morgenstern, U., Auer, S., Tramp, S., Martin, M.: Knowledge engineering for historians on the example of the catalogus professorum lipsiensis. In: International Semantic Web Conference. pp. 225–240. Springer (2010)
14. Tudorache, T., Nyulas, C.I., Noy, N.F., Musen, M.A.: Using semantic web in icd-11: three years down the road. In: International Semantic Web Conference. pp. 195–211. Springer (2013)