A Wiki-Oriented On-line Dictionary for Human and Social Sciences

Lydia Khelifa[1,2], Nadira Lammari[1], Hammou Fadili[3], Jacky Akoka[1]

[1] Conservatoire National des Arts et Métiers de Paris, 292 rue St Martin, 75141 Paris Cedex 03, France

[2] Ecole Nationale Supérieur d'Informatique d'Alger (ex INI), BP 68M Oued Smar, 16309, El Harrach, Alger, Algérie

[3] Fondation Maison des Sciences Humaine et Sociales de Paris, 54 boulevard Raspail, 75270 Paris Cedex 06, France

khelifalydia@gmail.com, {lammari, akoka}@cnam.fr, fadili@msh-paris.fr

Abstract. The aim of this paper is to contribute to the construction of a human and social sciences (HSS) on-line dictionary. The latter is Wiki-oriented. It takes into account the multicultural aspect of the HSS as well as the ISO 1951 international standard. This standard has been defined to harmonize the presentation of specialized/general and multilingual/monolingual dictionaries into a generic structure independent of the publishing media. The proposed Wiktionary will allow HSS researchers to exchange and to share their knowledge regardless of their geographical locations of work and/or of residence. After the conceptual description of this dictionary and the presentation of the mapping rules to Wiki semantic concepts, the paper will present an overview of the prototype that has been developed.

Keywords: Semantic Wiki, Human and Social Sciences, Multicultural Wiktionary.

1 Introduction

While social science studies human societies, human sciences deal with human groups and individuals, their history, their cultures, their accomplishments and their individual and social behaviors. Both social and human sciences (HSS) encompass heterogeneous disciplines like anthropology, sociology, economics, ethnology, geography, history, political science, archeology, linguistics science and religion science. They play a key role in understanding and interpreting the economic, cultural and social context of populations. The evolution of the research in this area inevitably involves knowledge exchange and sharing between researchers.

To promote exchanges between Maghrebi countries and France in the HSS area, the FMSH[1], with the collaboration of partners from France and Maghrebi countries[2], have defined a project aiming at the construction of a multicultural and multilingual content. This project will allow exchanges between Maghrebi and French researchers. It will also allow the sharing of knowledge related to the two cultures and to the two societies. In this project it has been decided to first construct an on-line dictionary for the HSS. This dictionary does not exist at the present time. It must respect the ISO 1951 standard [1], be extensible to many languages and exploit the Wiki technology. One of the reasons motivating the FMSH choice for the Wiki technology is the ease and the speed of defining, structuring and describing all types of data, according to different schema, using the WikiML (Wiki Markup Language). Moreover, the evolution management of this kind of application (dictionary application), generally difficult, is facilitated thanks to the Wiki platform, especially when the changes concern only the structure of the content.

The WikiMedia foundation supplies a Wiktionary. The latter is an open and universal dictionary. It is free for development and allows, authorized people to easily and rapidly edit, publish and maintain on-line content through collaborative processes that mutualize human skills. It also offers a complete versioning system and can alert anyone interested in particular themes when any content creation, modification or deletion, corresponding to his favorite themes, is performed. However, its current schema doesn't fulfill all the HSS dictionary functional requirements such as the search by context, hence, the idea to extend it.

The rest of the paper is organized as follows. Section 2 describes the peculiarities of the HSS on-line dictionary. Section 3 is dedicated to related works. Section 4 focuses on the conceptual modeling of this dictionary. The prototype is presented in Section 5. Section 6 concludes the paper and presents some perspectives.

2 The HSS dictionary description

To promote exchanges between of the two banks of the Mediterranean Sea in the HSS field, the development of a multilingual and multicultural e-dictionary has been initiated by the FMSH. This dictionary should, at first, contain the main HSS words used in France and in the Maghrebi countries, specify their use by both societies and supply their translation from one language to another one. This dictionary will be extended to all the languages of the Mediterranean countries later on.

The design of the on-line HSS dictionary must take into account the facts that:

• an entry A_k in a source language can have several meanings and therefore several translations B_1, …, B_m in the target language. Moreover, this same entry A_k can be defined with several components A_1, …, A_i of the dictionary schema (synonym, antonym, related nouns, pronunciation, etymology, etc). Each of these components could be an entry in the source language and could, therefore, have several meanings in the source language and several translations in the target language (Fig. 1). Let us

1 One of the acronym of the "La Fondation Maison des Sciences de l'Homme" (FMSH), http://www.fsp.maghreb-france.msh-paris.fr/

2 The partners are: FMSH, Cnam of Paris, ESI of Algiers.

note that any source language is also a target language. It depends on the required translation. Moreover, it may occur that an entry in the source language may not have a correspondent entry into a target language.

• the meaning assigned to a HSS dictionary entry depends on the context of the definition of this entry. The latter is described by a finite and known set of contextual parameters that vary from one discipline to another one. Among these parameters we can mention geographic and temporal parameters for sociology.

• the components used for the description of an entry are those of the ISO 1951 standard [1].

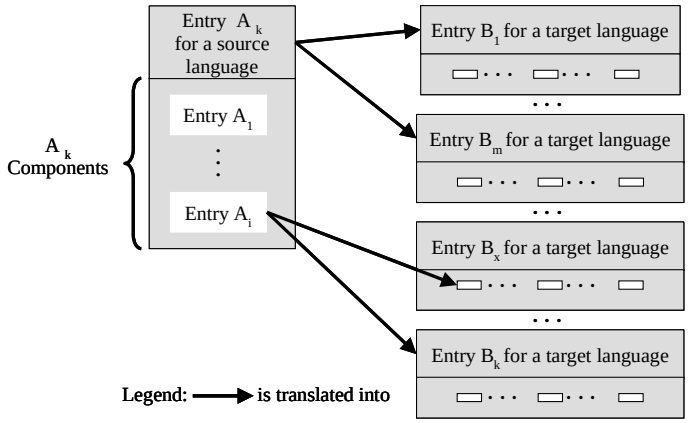

Fig. 1. The HSS dictionary schema extract.

Beside the constraints related to the description of the HSS one-line dictionary, its development, were in the specification document, conditioned by the exploitation of the Wiki technology for its advantages including the ease of construction and maintenance of collaborative contents by non expert users (users who are not specialists in computer science). Finally, it must allow the search by context.

3 Related Works

There are several projects for the construction of specialized on-line dictionaries. Among them, we can mention the PAPILLON project [2], the DHYDRO project [3], the JMdict/EDICT project [4] and the SAIKAM project [5]. In the PAPILLON project, the paradigm of the Linux collaborative building has been applied to the collaborative edition of definitions. It offers, among possible search criteria, the retrieving of a word according to its contextual reading. In the DHYDRO project a terminological and multilingual space specialized for the hydrographic domain has been built. JMdict/EDICT proposes a remote edition tool for a multilingual terminological database. SAIKAM is an on-line dictionary. It aims at the creation of new Thai words for Japanese ones.

Let us note also that the Semantic Web Deployment Working Group, part of the W3C Semantic Web Activity, recommended, since august 2009, the SKOS (Simple

Knowledge Organization System) model for the description of thesaurus, taxonomies or any other controlled vocabularies [11]. SKOS is based on the RDFS language.

However, none of the projects cited above uses the Wiki technology. This led us to explore the possibility to exploit the current Wiktionary project of WIKIMEDIA foundation. The latter proposes a Wiktionary per language. Some of them, like the Arabic Wiktionary, lacks structure. The other ones don't have similar structure (eg. English and French Wiktionary [13] [14]). For example, the French Wiktionary is organized into articles [13]. Each article is used to describe a word. It gathers:

- a main section for the description of the word in the language associated to the Wiktionary,
- zero or more other language sections, each for a language different from that of the Wiktionary,
- a categorization section that classifies the word into one or more categories from those listed
- and finally, a section that allows to establish links between the article and other ones in the others Wiktionaries. These links are oriented to articles having the same title. They don't concern their translations.

The main section proposes:

- a mandatory set of basic description elements: etymology, one or more sections for the type of word (i.e its spelling variants, its abbreviation, its derived words, its synonyms, its hyponyms, its translations, etc.)
- and a set of optional elements: pronunciations, anagrams, and a section «to see also» that gather the links related to the article and a reference section that gives the references used during the edition of the article.

The sections dedicated to languages are similar to the main section except that it doesn't contain some sections like the one needed for translation or for hyponymy.

The description possibilities supplied by the current Wiktionary project don't meet the HSS dictionary specificities. On one hand, it lacks an automatic management of correspondences that allows managing the complexity of referrals between the source language and the target language. It is possible to use the current Wiktionary to change an entry regardless of other entries to which it is linked. In other words, it is possible to add in a Wiktionary dedicated to one language A, a translation of a word into a language B without impacting the change in a Wiktionary dedicated to language B. Moreover, links between Wikis, in the Wiktionary, can be established only between articles having the same name. This means that we can not link two words, such that the first one is the translation of the second one, if the two words are not in the same Wiki. On the other hand, the current Wiktionary project does not allow contextual search of the meaning of words. This functionality is very important in HSS field and must be fulfilled by the HSS Wiktionary application.

Another version of a Wiktionary exists: OmegaWiki [12]. It is based on an extension of MediaWiki. OmegaWiki unlike the current Wiktionary project gathers in one space all the Wiktionaries. It overcomes the drawback of the current Wiktionary concerning the impact of changes from one Wiktionary to another. Finally, OmegaWiki, at the present time, can be used only for search and it does not supply a contextual search for the word meaning.

4 The HSS Wiktionary Design Approach

As mentioned in the previous section, a HSS on-line dictionary entry could have many descriptions. Each of these descriptions can be valid for a given context described by a set of contextual parameters like geographic and temporal parameters. Moreover, each description must respect the ISO 1951 standard. The design of the HSS on-line dictionary is, therefore, based on the correspondence between an entry and its contexts of definition in a source language and an entry and its contexts of definition in the target language. This correspondence is performed according to a schema that could contain the definition of the entry, the synonyms, the antonyms, the related words, the pronunciation, the spelling, etc.

The conceptual description of such dictionary could be represented using an UML class diagram. Figure 2 is an extract of this conceptual model. This model shows that the description of an HSS dictionary entry (word) in a given language is obtained by gathering the variants of this description. Each variant of a description corresponds to a context defined by the concerned discipline, the set of context elements which are context parameters values. Each discipline has its own context parameters. Each entry described with a given variant of a description could have a synonym related to this variant of description.

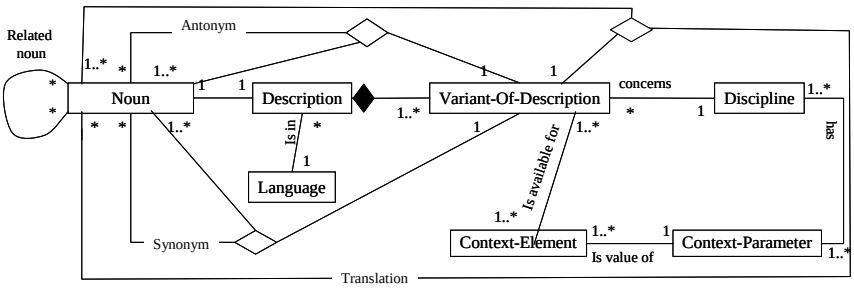

Fig. 2. An extract of the HSS dictionary conceptual model.

The use of the Wiki technology for HSS on-line dictionary constitutes, in the project, a technical constraint that we must comply to. To date, there are several Wikis. WikiNi, Wiclear, DokuWiki, MediaWiki and semantic Wikis are some examples of existing ones. Semantic Wikis such as KawaWiki [6], IkeWiki [7], SweetWiki [8], Kaukolu [9] and Semantic MediaWiki [10] are semantic web extensions of the Wikis. KawaWiki allows the creation of Wiki pages using RDF templates and their querying by means of SPARQL. IkeWiki is a tool for formalized and collaborative building of content. It offers the possibility to annotate the links and the possibility of reasoning. SweetWiki semantically annotates Wiki resources. It supports the social tagging, uses ontologies for the structures of the Wiki and offers a WYSIWYG editor. Kaukolu is a semantic Wiki based on JSPWiki. It allows the annotation, creation and display of pages. It also replaces Unified Resource Identifiers by alias to allow creation of new pages. Semantic MediaWiki is an extension of MediaWiki. It inherits the advantages of MediaWiki such as easiness to use, editing collaborative documents (minimum of technical prerequisite), and its evolution. It

also allows annotating Wiki pages, their content and the links between them. Moreover, for navigation purposes, the semantic Wikis, and in general the Wikis, allow the intensive use of hyperlinks. Therefore, a future user of the application can get a global view of a page and can then have a zoom (a detail) of the part of the content he (or she) is interested in.

Our study of the state of the art and its confrontation with HSS on-line dictionary peculiarities, allows us to retain, for its realization, the Semantic MediaWiki technology.

The concepts associated with a Semantic MediaWiki are represented in the metamodel of Figure 3. A Semantic MediaWiki, as shown in Figure 3, is a set of Wiki pages that can be annotated. A Wiki page can be related to another one through external hyperlinks. Hyperlinks can also be used within a page. They can also be annotated.

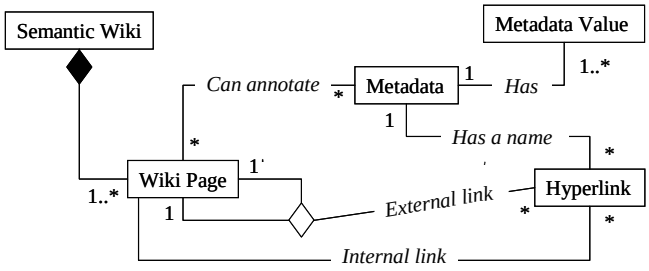

Fig. 3. The Semantic MediaWiki metamodel.

The mapping between the concepts of our on-line dictionary (Fig. 2) and the concepts of the Semantic MediaWiki (Fig. 3) is presented in Table 1.

Table 1. The mapping between HSS Wiktionary concepts and Semantic MediaWiki concepts.

HSS on-line dictionary concepts	Semantic MediaWiki concepts
Description/Variant of description	Wiki page
Context element	Metadata value of a context parameter
Language/ Discipline / Context parameter	Metadata
Antonym/Synonym/Related noun/Translation	Hyperlink

This table shows that the different descriptions of an entry (variant of a description) are mapped, in a Semantic MediaWiki, to Wiki pages. This is the same for the complete description of an entry. The concepts “Language”, “Discipline” and “Context parameter” are metadata. A context element of the HSS dictionary is mapped into a metadata value that can be taken by its corresponding context parameter. All the other concepts (Antonym, Related noun, Synonym and Translations) are translated into Wiki links.

Finally, to insure the extensibility of our Wiktionary to many languages (such as the Amazigh) and dialects of the Maghrebi countries, we propose to build a Wiki per language. The example of Figure 4 illustrates the structure of our HSS Wiktionary. This figure describes a French Wiki page for a variant of description of the word

"Entrepreneur" (one of its meanings in English is "contractor"). This page is annotated by the following metadata values:
• "Entrepreneur" is associated with the metadata "Word",
• "Sociologie" which corresponds to a value of the metadata "Discipline",
• "Français" which corresponds to the value of the metadata "Language",
• "13ième siècle" and "Maghreb" are respectively values of metadata temporal and geographic parameter. These two parameters represent the context elements of the context parameter "Discipline".

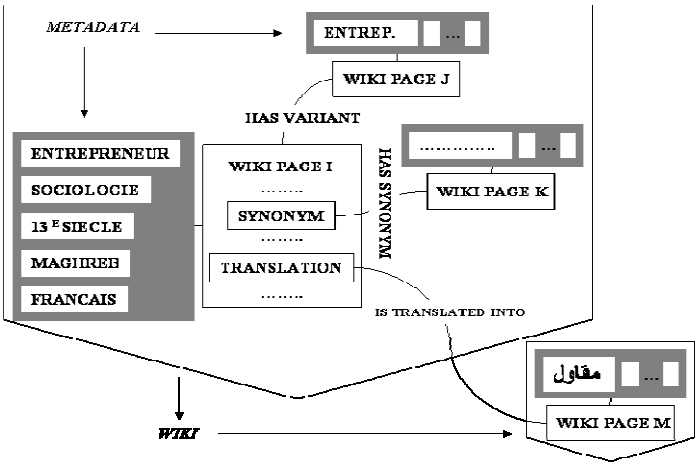

Fig. 4. An illustration of the HSS Wiktionary structure through an example.

The Wiki page associated to this variant of description of the word "Entrepreneur" is linked, in Figure 4, to other variants via the hyperlink "has variant". Moreover, this variant of description of the word "Entrepreneur" contains a hyperlink "is translated into" that links this Wiki page to the Wiki page representing the translation into Arabic of the word "Entrepreneur" for the same context of definition.

5 The Prototype

After translating the conceptual schema of our dictionary into a logical schema respecting the Semantic MediaWiki technology, we built a HSS Wiktionary prototype. Thus, we have chosen to build a Wiki by language and to establish links between them. Such a choice, allows us to construct a French-Arabic Wiktionary and then to extend it to other languages and dialects of the Mediterranean countries. Figure 5 is the welcome page of the Wiktionary. Through this page the user can enter a kind of Wiktionary (at the present time French and Arabic ones). He can also ask for the definition of a word (or its synonyms, or its close words) by giving all or some information (values of the context elements) about the context.

Fig . 5. The welcome page of our Wiktionary application.

The editor of the HSS Wiktionary (Fig. 6) includes, at the present time, a subset of the elements of ISO 1951 standard. Its extension to all elements of this standard or only to those useful for HSS field is possible. Using this editor, the user could annotate a Wiki page associated to an entry, by the metadata of its context of definition. He could also complete its description by using annotations associated with the elements of the schema issued from the ISO 1951 standard. Before entering a description (in a given language) of an entry (word) the user must first provide the context of the definition of this entry (i.e. the user must enter the discipline, the language concerned by the entry and the other context elements that validate and specialize its description). According to the context provided, the system will either propose to modify the last version of the description (if the entry already exists with the same context) or to create it. During the modification of an existing description (page) or its creation, the user has to use the proposed tags to add possible synonyms, antonyms, related nouns of the entry. Semantic MediaWiki translates these metadata into RDF.

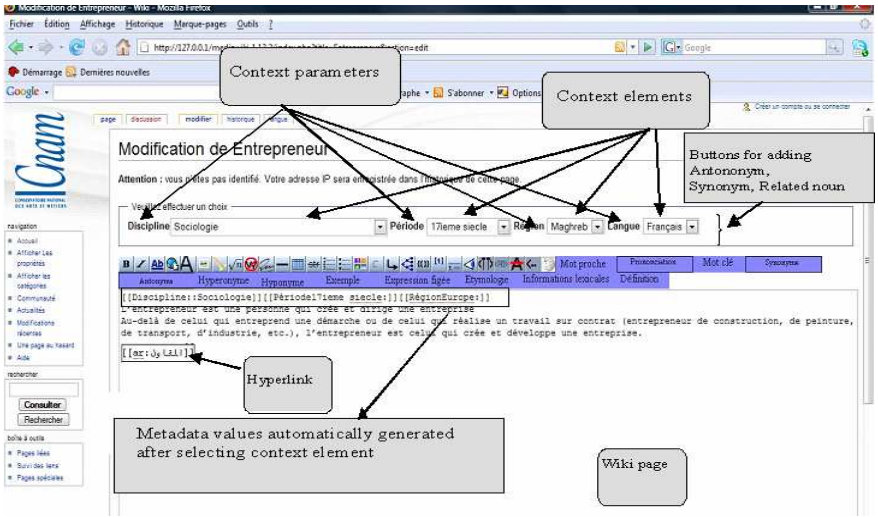

Fig. 6. Editor interface of the French Wiktionary.

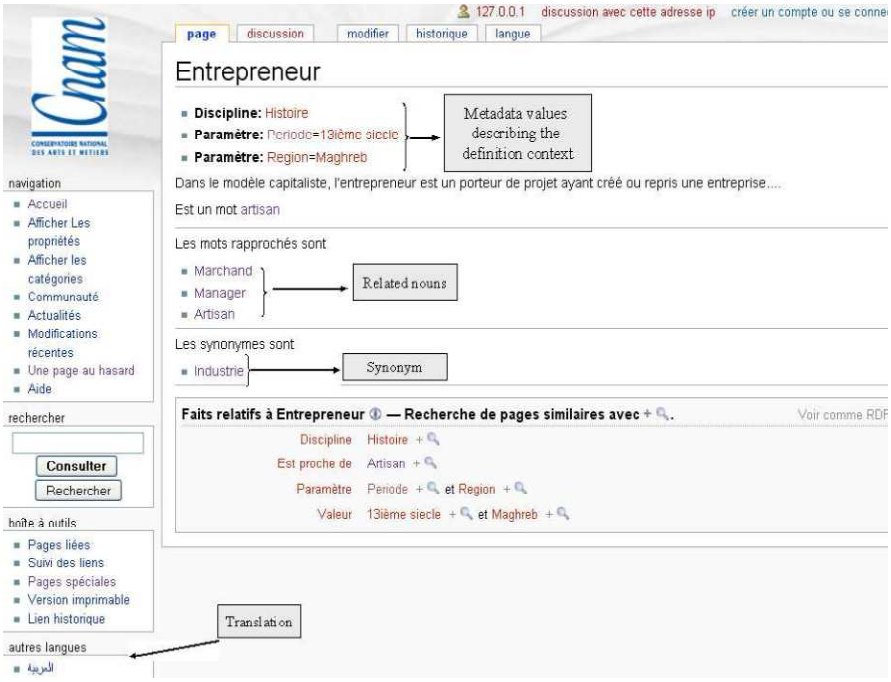

Fig. 7. An example of Wiki page consultation.

Note that due to the multicultural aspects of the HSS Wiktionary, an entry may not have a correspondent in a given target language. Let us note also that the global description of an entry could be obtained by gathering its variants in a Wiki page. The user may also wish to consult a description of an entry for a given context. The system, in this case, provides the description in which hyperlinks to synonyms, antonyms, related words and a correspondent translation appear. For example, the interface of Figure 7 is provided to a user who wants to obtain a description of the French word “Entrepreneur” for the context described by the metadata values.

6 Conclusion

We have described in this paper the HSS on-line dictionary. For this purpose, we have used, as required in the project specification document, the Wiki technology. The latter makes the content editable collaboratively and facilitates its exploitation.

After presenting the specificities of our on-line dictionary, we have synthesized them using an UML conceptual model. By taking into account the technical constraint associated to its implementation, we have proposed a first version of a prototype resulting from the mapping between the conceptual model of the dictionary and the Semantic MediaWiki metamodel. To evaluate the success of this first version, we asked experts, from different disciplines, to populate it with HSS words. While populating it, a cultural exchange of knowledge between researchers from the

Mediterranean countries will take place, allowing to share this knowledge between society members.
Future research will tackle the issue related to the separation between the presentation and the storage layer of the dictionary. As of today this separation was not possible for time reason. We intend to take advantage of SKOS (the W3C recommendation for the representation of thesaurus, taxonomies or any other controlled vocabularies) to perform such a separation. The latter will lead us to map the dictionary into multiple formats. In addition, we will take into account the access management aspect related to the security issues of the Wiktionary application. Finally, we will integrate the Amazigh language and its graphical symbols into the Wiktionary.

References

1. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=36609
2. Mathieu Mangeot: Papillon project: Retrospective and Perspectives. Proc. of Acquiring and Representing Multilingual, Specialized Lexicons: the Case of Biomedicine, Ed. Pierre Zweigenbaum. LREC workshop 2006, Genoa, Italy, 22 May 2006.
3. Sylviane Descotte, Jean Luc Husson, L. Romary, M. Van Campenhoudt and N. Viscogliosi: From specialised lexicography to conceptual databases: which format for a multilingual maritime dictionary. The 2d International Conference on Maritime Terminology. Turku, Finland. 12 May 1999.
4. Francis Bond, Jim Breen: Semi-automatic refinement of the JMdict/EDICT Japanese-English dictionary. 13th Annual Meeting of The Association for Natural Language Processing, pages 364-367, Kyoto, 2007.
5. Vuthichai Ampornaramveth, Akiko Aizawa, Saikam: Collaborative japanese-thai dictionary development on the internet. The Asian Association for Lexicography (ASIALEX) Biennial Conference, Korea, 2001.
6. Kensaku Kawamoto, Yasuhiko Kitamura and Yuri Tijerino, KawaWiki: A SemanticWiki Based on RDF Templates. Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops WI-IATW'06),2006.
7. Sebastian Schaffert, IkeWiki: A Semantic Wiki for Collaborative Knowledge Management. 15th IEEE International Workshops on STICA06, Manchester, UK, June 2006.
8. Michel Buffa, Gaël Crova, Fabien Gandon, Claire Lecompte, Jeremy Passeron, SweetWiki: A semantic Wiki. Special Issue of the Journal of Web Semantics on Semantic Web and Web 2.0, Volume 6, Issue 1, Pages 84-89, February 2008.
9. Malte Kiesel: Kaukolu: Hub of the Semantic Corporate Intranet. Workshop From Wiki to Semantics, ESWC 2006.
10. Markus Krötzsch, Denny Vrandecic, and Max Völkel: Semantic MediaWiki. The 5th International Semantic Web Conference, ISWC2006, 2006
11. SKOS. http://www.w3.org/2004/02/skos/
12. OmegaWiki. http://www.omegaWiki.org/Meta:Main_Page
13. Wiktionary: English Entry Layout http://en.wiktionary.org/wiki/Wiktionary:Entry_layout_explained
14. Wiktionary: French Entry Layout. http://fr.wiktionary.org/wiki/Wiktionnaire:Structure_des_articles