Dictionary Building with the Jibiki Platform

Mathieu Mangeot
Condillac – LISTIC – Université de Savoie
Campus Scientifique
F-73376 LE BOURGET DU LAC CEDEX
Mathieu.Mangeot@univ-savoie.fr

Abstract
The Jibiki platform is an online generic environment for writing and querying all kinds of dictionaries: terminological glossaries, bilingual dictionaries, multilingual lexical databases, etc. It has been developed mainly by Mathieu Mangeot (Université de Savoie, France) and Gilles Sérasset (Université de Grenoble 1, France), thanks to research driven by the GETA team of the CLIPS laboratory in Grenoble, France. The platform allows one to lookup all the dictionaries available on the server and to display the results in the same window. The advanced query interface offers a combination of multiple search criteria. The writing of the entries is done directly online on the platform via a web browser. The writing interface is generated automatically from the description of the structure of the entries (an XML schema), thus allowing the edition of (almost) any type of dictionary entry.

1 Overview of the platform

The Jibiki platform is an online generic environment for writing and querying all kinds of dictionaries: terminological glossaries, bilingual dictionaries, multilingual lexical databases, etc. It has been developed mainly by Mathieu Mangeot (Université de Savoie, France) and Gilles Sérasset (Université de Grenoble 1, France), now helped by Francis Brunet-Manquat, thanks to research driven by the GETA team of the CLIPS laboratory in Grenoble, France.

The platform is implemented in Java, exclusively with open source tools. It is based on Enhydra, a web server of dynamic java objects and Postgres, a relational database. The interface is available in English, Estonian, French, German and Japanese. New languages can be easily added. Annex tools have been added on various instances of the platform. Some facilitate the communication between communities of users (forums, distribution lists) and others, the work of the lexicographers (tool for managing aligned bilingual corpora).

2 Comparison with existing software

In this section, we will briefly compare our software with two other well known dictionary building software: TshwaneLex and IDM's Dictionary Publishing Software.

	Jibiki platform	Tshwanelex	IDM DPS
Price	Not for commercial Free for academic	1900 € for commercial 150 € for academic	?

Online searching	Yes	Yes	Unclear
Online editing	Yes	No	?
Unicode handling	Yes	Yes	Yes
XML structure	Yes	Unclear	Yes
Importing existing XML dictionary	Yes	No	No
Corpus handling	Not included	Not included	Included

3 Projects currently using the platform

The platform is currently used by three lexicographical or terminological projects.

3.1 Papillon Project

This project,[1] launched in 2001, is at the origin of the building of the platform. Its main goal is the construction of a multilingual lexical database with a pivot structure covering among others the following languages: Chinese, English, French, German, Japanese, Lao, Malay, Thai and Vietnamese. The resulting resources are publicly available and free of rights. The projects is open to all those who are interested in these languages.

3.2 GDEF Project

The GDEF[2] (Great Estonian-French Dictionary) started in 2003. Its goal is to build a bilingual Estonian-French dictionary of about 80,000 entries, by a team of 8 people made of linguists, as well as Estonian and French translators.

3.3 LexALP Project

The European project LexALP, launched in 2005, use the Papillon platform in order to develop a terminological database of legal and administrative terms in the main Alpine languages (French, German, Italian and Slovene).

4 Dictionary lookup

The platform allows one to lookup all the dictionaries available on the server and to display the results in the same window. The advanced query interface offers a combination of multiple search criteria on:

- the languages: source, targets, available resources;
- the character string: prefix, suffix, substring;
- the content of the entries: headword, variants, pronunciation, domain, gloss, part-of-speech, translations, examples, etc.

It is even possible to define new search criteria when a new resource is added by defining common pointers on searchable information parts. In the case where a normal search returned no results, a reverse lookup is also executed.

[1] http://www.papillon-dictionary.org
[2] http://estfra.ee

5 Entries writing

The writing of the entries is done directly online on the platform via a web browser. The writing interface is generated automatically from the description of the structure of the entries (an XML schema), thus allowing the edition of (almost) any type of dictionary entry.

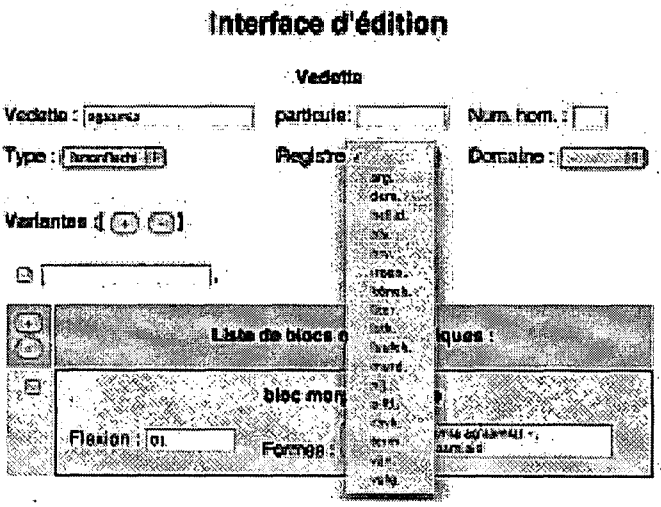
The interface, built upon an HTML form can also deal with relatively complex structures thanks to more complex interactors that combine the basic HTML ones (text boxes, radio buttons, pop-up menus). Such example is the list management one that allows the writers to add, delete or reorder elements in a list by simply clicking on a button. These elements can be themselves complex objects containing lists of other objects, etc. A specific module allows the writer to establish links to entries in other resources available on the server. This technique is mainly used for linking an entry to its translation in another language when the translation already exist as a separate entry. The writing process is divided in several steps depending on the project. The GDEF is the most complete with three steps:

1) A contributor writes an entry;

2) It is next revised by a reviewer;

3) It is then validated by a validator;

When the entry is validated, it is integrated into the dictionary and all the users can search it.



6 Tasks management

In order to manage the different tasks and roles, the platform gives the possibility to define groups and access rights. There are several groups with predefined rights:

- If the user is not logged, it can lookup the public resources available on the platform.
- When the user is registered and logged, s/he is included de facto in the *contributors* group and can contribute through the entry writing interface.

• The users members of the *reviewers* group can revise the contributions written by users members of their working group.

• The users members of the *validators* group can validate the previously revised contributions.

• Then, the server *administrators* manage users and their groups, add new resources, etc.

In order to facilitate the construction work of the dictionary and possibly the remuneration of the writers, it is possible to obtain a summary of all the contributions in a given period of time. Then, the dictionaries can be exported as a whole or by parts, in several formats (text, HTML, XML, PDF, etc.) and printed.

References

Chalvin, A., Mangeot, M. (2006), 'Méthodes et outils pour la lexicographie bilingue en ligne: le cas du grand dictionnaire estonien-français', in *Proc. EURALEX 2006*, à paraître, Turin, Italie, 6-9 septembre.

Mangeot, M. (2001), *Environnements centralisés et distribués pour lexicographes et lexicologues en contexte multilingue*. Thèse de doctorat, Spécialité Informatique, Université Joseph Fourier Grenoble I.

Mangeot, M., Sérasset, G., Lafourcade, M. (2003), 'Construction collaborative de données lexicales multilingues, le projet Papillon. Les dictionnaires électroniques: pour les personnes, les machines ou pour les deux?', in Zock, M., Carroll, J. (ed.), *TAL Traitement Automatique des Langues*, Vol. 44: 2/2003, pp. 151-176.

Mangeot, M., Thevenin, D. (2004), 'Online generic editing of heterogeneous dictionary entries in Papillon project', in *Proc. of the COLING 2004 conference*, vol. 2, Geneva, Switzerland, 26 August, pp. 1029-1035.

Sérasset, G. (2004), 'A generic collaborative platform for multilingual lexical database development', in Sérasset, G. (ed.), *COLING 2004 Multilingual Linguistic Resources Workshop*, Geneva, Switzerland, 28 August, pp. 73-79.

Gilles Sérasset (2005), 'Multilingual legal terminology on the jibiki platform: The lexalp project', in Lafourcade M. (ed.), *Proc. of Papillon 2005 Workshop*, Chiang Rai, Thailand, 11-13 December, pp. 64-73.