



HAL
open science

Current Status of the Papillon Server Development

Gilles Sérasset

► **To cite this version:**

Gilles Sérasset. Current Status of the Papillon Server Development. Papillon 2002 Seminar, Jul 2002, Tokyo, Japan. pp.7. hal-00965854

HAL Id: hal-00965854

<https://hal.science/hal-00965854>

Submitted on 25 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Current Status of the Papillon Server Development

Gilles SÉRASSET

GETA, CLIPS, IMAG, 385 rue de la Bibliothèque, BP 53
38041 Grenoble cedex 9, France
Gilles.Serasset@imag.fr

Abstract

This paper gives an overview of the current Papillon web server and the services it provides.

Keywords: Papillon multilingual database, dynamic web server, Enhydra

Résumé

Cet article donne un aperçu du site web Papillon et des services fournis.

Mots-clés Base de données lexicale multilingue Papillon, serveur web dynamique, Enhydra

Introduction

One of the main Papillon project's objectives is to provide a web server where Papillon contributor will find different services. After an intensive work, the developers (mainly Mathieu Mangeot-Lerebours and Gilles Sérasset) have finally achieved the core of this dynamic web server.

Most of the services are now unimplemented and the implemented ones are currently in a beta version, but we finally succeeded in opening the first version of the Papillon web server. This server is accessible through the URL: <http://www.papillon-dictionary.org/>.

This paper presents the current state of this server. We will first give an external overview of the server, then we will discuss and present its implementation. We will describe the implemented services give some perspectives on the services that are planned.

1 External view of the Papillon web server

When entering the Papillon web site, the user is welcomed by a page summarizing the main objectives of the Papillon project. Like every other page of the Papillon web site, this page is composed of 3 parts:

A header, used to navigate inside the main sections of the web site (Part I). All pages share this header.

A menu, that provides a quick access to different links (Part II). This menu is always present, but the links may vary depending on the section or the user.

A content (Part III) that is specific to the page.



1.1 “Information” section

The information section is the central place where any user should find any information concerning the Papillon project, from its objectives to its current implementation. It currently offers 2 services:

- A mailing list archive.
- A document repository.

The mailing list archive provides a searchable archive of all mails exchanged by Ppillon partners through the official papillon mailing list. The Information repository offers a searchable repository of different documents providing information about the Papillon project. Here, the user can find different documents containing:

- Some general information about the papillon project
- Published articles related to the Papillon project
- Information about the development of Papillon’s server and tools
- The description of the different dictionaries available on the server
- Some simple “How To” recipes on tasks involved by the Papillon project
- Other miscellaneous information

These services are detailed in section 3.

1.2 “Consultation” section

The consultation section is the place where the user has access to the lexical data available on the Papillon server. This section currently offers a unique general interface giving access to all the dictionaries (as in [2]).

Source: English	Targets: ANY English French Japanese	Resources: ANY FeM JMDict Papillon
Lemmatize the entry: <input type="checkbox"/>	Headword: <input type="text"/>	Pos: <input type="text"/>
Any part: <input type="text"/>		
Go		

This integrated interface gives access to the dictionaries belonging to the “Purgatory layer” (as defined in [3]) as well as to the Papillon multilingual dictionary itself (Paradise layer, see [13, 14]).

1.2.1 Purgatory layer

This interface will require more development as the general interface is not informative enough to be used by general users (number of entries available for each dictionary, macrostructure of each dictionary...). However, it is currently functional. Hence, if I am looking at the Headword “meurtre” (murder) in the French Vietnamese dictionary, I’ll get the following result:

Source: French	Targets: ANY English French Japanese	Resources: JMDICT Papillon ThaiDict VietDict
Lemmatize the entry: <input type="checkbox"/>	Headword: meurtre	Pos: <input type="text"/>
Any part: <input type="text"/>		
Go		

[FeM](#) [Papillon](#) [XML](#) [DEFAULT](#) [DEC](#) [JMDict](#) [ThaiDict](#) [VietDict](#)
meurtre

danh từ giống đực sự giết người *crier au meurtre* kêu ca ãm ã

This output is a formatted view of the data stored in the Papillon web site. This view may be customized easily, however, in the current implementation, this customization cannot be done directly by a user, as the customization interface is not implemented yet.

Several views are available for each dictionary. Among these views, the XML view faithfully reflects the structure of each dictionary (every dictionary is stored in XML format).

Source: French	Targets: ANY English French Japanese	Resources: JMDict Papillon ThaiDict VietDict
Lemma: meurtre	Pos:	Any part:
Go		

FeM Papillon XML DEFAULT DEC JMDict ThaiDict VietDict

```

<entry>
  <headword>meurtre</headword>
  <syntactic-sense>
    <pos>danh từ giống đực</pos>
    <translation lang="vie">sự giết người</translation>
    <example>
      <fra>crier au meurtre</fra>
      <vie>kêu ca ầm ĩ</vie>
    </example>
  </syntactic-sense>
</entry>

```

1.2.2 Paradise layer

Asking for the same headword (“meurtre”) in the Papillon dictionary will lead to the following result (see [4, 5, 6]). As see above, the result is presented in a customizable form. But, when dealing with entries of the Paradise layer (of the Papillon dictionary per se), we are also able to present entries from other languages that correspond to the same interlingual acception as defined in [7, 8] and prototyped in [1] (in our example, in Japanese and Malay).

MEURTRE , nom, masc

action de tuer: ~ PAR L'individu X DE L'individu YQSyn : assassinat, homicide#1, crimelV0 : tuertA0 : meurtrier-adj IS1 : auteur [de ART ~] // meurtrier-n IS2 : victime [de ART ~] IMagn : atroce, affreux, brutal, horrible, inqualifiable, odieux IA2Prepar1 : avec préméditation, prémédité | postpos // assassinatITel qu'il y a deux/trois/quatre Y : double/triple/quadruple | antépos ["Les victimes de ce double meurtre sont un père de famille et son fils de 15 ans."] IOper1 : accomplir, commettre, perpétrer [ART ~] | tremper [dans ART ~] ["Il a refusé de tremper dans ce meurtre odieux."] ICausOper1 : pousser [N=X au ~] IS1CausOper1 : mobile [de ART ~] IReal?--I : enquêter [sur ART ~] IReal?--II : élucider [ART ~], trouver l'auteur de ART ~IReal?--III : punir, châtier [ART ~] | venger [ART ~] IS1Real?--III : brigade criminelle IPrepar1 : préméditer, préparer [ART ~] // comploterIEssayer de faire croire qu'un M. est un N : maquiller [ART ~ en N] ["Il a maquillé ce meurtre en accident/suicide."] IFait de tenter un M. : tentative [de ~] ICri lancé par qqn. qui assiste à <risque d'être victime d'un M. : « Au ~ ! » IC'est ici que le double meurtre a été commis. Soupçonné du meurtre de son épouse, il a été arrêté par les gendarmes mercredi. Il devrait comparaître aux assises dans trois semaines comme auteur présumé du meurtre d'un quinquagénaire.-----La mésestante pourrait être le mobile du meurtre._appel au meurtre_crier au meurtre_

殺人, 名詞 /satujin/

dousa hito X NO hito Y Qsyn : 殺戮, 殺害 IConv : 殺傷, 暗殺 IMagn : ひどい(hidoi), むごい(mugoi) IQS1 : 殺人者(satsujinsya) 殺人鬼(satsujinki) IQS2 : 被害者(higaisya) IOper1 : [- wo] する(suru) [- wo] 犯す(okasu) IPrepar : 計画殺人(keikakusatujin), 殺人未遂(satujinmisui), 予謀殺人(yobōsatujin) IInstr : 凶器(kyouki) IA0 : 殺人的(satujinteki) IResult : 殺人罪(satujinzai), 殺人犯(satujinhan) IInvol : 容疑者(satujinhan) I殺人を犯す(commit murder)殺人剣(satujinken)殺人光線(satujinkōsen)囑託殺人(syokutakusatujin)

bunuh, n. /pem+bu+nu+han/

perbuatan membunuh: ~ OLEH L' individu X KE ATAS L' individu Y Qsyn : homisid IV0 : membunuh-v.tr. IS1 : (Nama bagi X) pelaku : pembunuhan-n IS2 : (Nama bagi Y) mangsa : bunuhan-n IMagn : mengerikan, kejam IOper1 : (Melakukan P.) melakukan Itdk pernah berlaku pembunuhan di daerah kami .polis menjumpai senjata yg digunakan dlm pembunuhan itu .dia didakwa melakukan pembunuhan darjah pertama .

1.3 “Edition” section

The “Edition” section is the place where the user will be able to contribute to the Papillon dictionary. He will be able to correct existing entries and add new ones in the dictionary. This section is not implemented yet as it requires a lot of development (management of private user spaces, interface to the Papillon dictionary’s microstructure...).

Until now, users willing to contribute entries to the dictionary may use his favorite XML editor to produce a valid Papillon entry. Then send it to the Papillon web site maintainer who will add it to the web site. You should send an email to the Papillon web site maintainer to get an example of a Papillon entry.

1.4 “Contacts” section

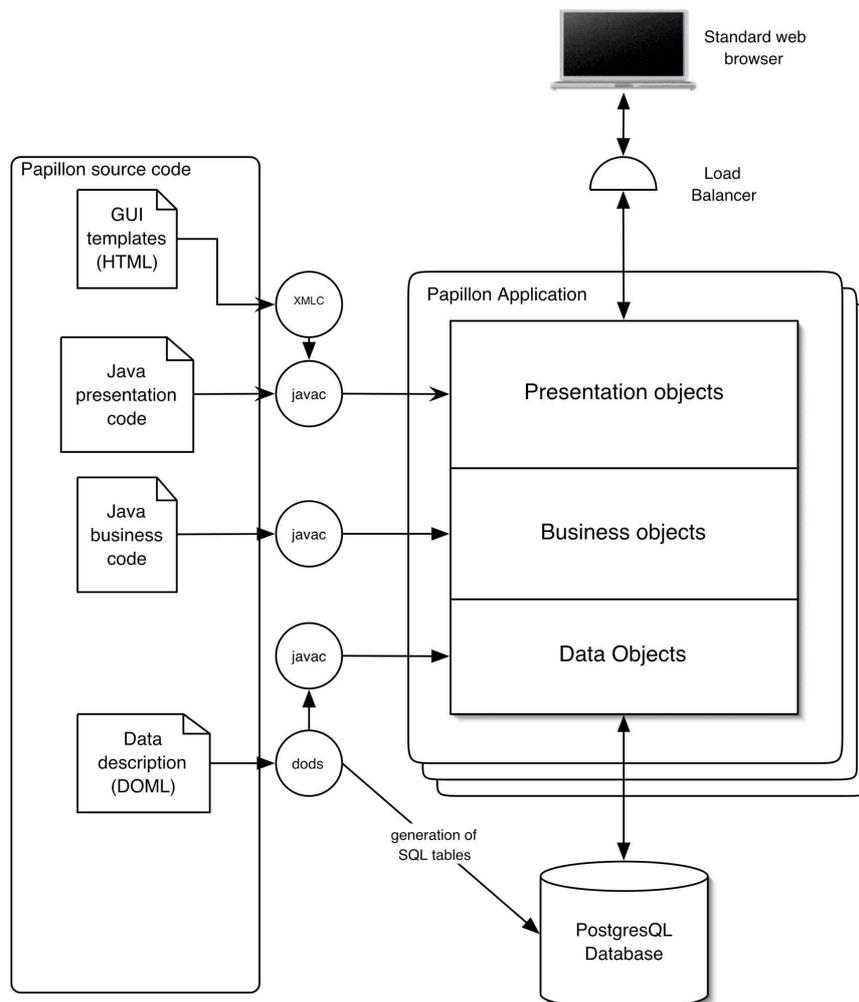
This section gives addresses of persons in charge of the Papillon project.

1.5 “Help” section

This section will contain documents describing the Papillon web site and giving simple help on the web site organization.

2 Software architecture

Papillon web site is a dynamic web site built with the “Enhydra” framework (an open source Java Application Server). This framework provides different tools and managers that are extensively used by the Papillon application.



The Papillon application features a 3-tier architecture:

The **presentation** layer contains code that specifies the graphical user interface (GUI) of this application. In our context, this GUI is implemented through a standard web browser. Hence, contains HTML templates and JAVA classes that implement the GUI behavior.

The **business** layer contains the classes that are specific to the Papillon application.

The **data** layer contains the persistent classes and the methods that perform the database requests.

With this architecture, it will be easy to adapt the Papillon application to a different context. To adapt it to be used on cellular phone, we will only modify the presentation layer and provide HTML data (for I-mode phones) or WML file for WAP phones. It will also be relatively easy to adapt the data layer to a XML database.

All requests to the Papillon application are served through a load balancer that distributes requests among different machines. Hence, it will be easy to handle more and more requests without involving any additional development.

The Enhydra framework allows a clear separation between the presentation code (that implements the behavior of the GUI) and the HTML templates (that specify the appearance of the GUI). Hence, the appearance may be modified by graphic designers with no particular knowledge of any programming language (unlike PHP or ASP pages).

3 Implemented services

Among the services currently provided by the Papillon application, the user will find an information repository and a mailing list archive.

3.1 Information documents repository

The Papillon application is organized to allow contributive work from Internet users. In our context, users are not only able to provide lexical data. They are also able to provide additional information that has to be integrated to the Papillon web site.

For that purpose, users having the sufficient privileges will be able to easily upload documents that will be integrated to the Papillon web site. When uploading such a document, users will provide its title, its author and the section in which he want it to appear (for a list of available sections, see paragraph 1.1).

In this repository, we accept PDF documents, DOC or RTF (word) files, GIF or JPG images or a simple HTML file. In addition to these kinds of files, we also accept a ZIP or TAR archive containing a whole HTML document (consisting in different HTML and GIF/JPG files linked together) that will be integrated to the site.

When submitting a HTML document, the user only has to focus on its content. The Papillon server will automatically add the header and menu composing the standard Papillon layout. The following figure shows such a HTML document, as seen by any Papillon user after its addition in the repository.



The screenshot shows the Papillon website interface. At the top left is a butterfly logo next to the word "Papillon". A navigation menu includes "Information", "Consultation", "Edition", "Contacts", and "Help". On the left side, there is a sidebar with links: "Register", "Login", "Clips", "NII", "Old Site", "English", and "français". The main content area features a central announcement for the "Atelier PAPILLON-2002 Workshop sur les bases lexicales multilingues on Multilingual Lexical Databases" in Tokyo, Japan, from July 16-18, 2002. Below this, a red banner states "Date limite de soumission étendue au 31 mai 2002" and "Submission deadline extended to the 31 May 2002". A link for "Présentation" is provided. At the bottom, it says "ENGLISH VERSION BELOW".

3.2 Mailing list archive

The Papillon project currently maintains a mailing list gathering all participants. This mailing list is a very valuable resource for information. Hence an archive of this mailing list is really useful.

However, the Papillon mailing list is particular as it gathers participants from a lot of countries speaking about languages. This fact leads to mails that use different encodings. None of the currently available mailing list tools we know are able to correctly handle such different encodings. Hence, we had to patch one of them (MhonArc) to handle these encodings.

Mails sent to the mailing list are converted to UTF-8, then uploaded into the Papillon database where they are immediately available to be viewed by Papillon web site users.

This archive is searchable.

Conclusion and perspectives

The first beta release of the Papillon application is now available on the web. It provides some of the services that are planned by the Papillon project. This release is a good starting point for the future developments of the Papillon project.

We chose to first implement the tools that will allow Papillon partners to communicate easily. With a mailing list archive and a document repository the partners will be able to complete the current phase that involves a lot of specification and training. Moreover, the “Purgatory” layer (as defined in Mathieu Mangeot’s PhD Thesis) is implemented, allowing users to get an integrated access to many dictionaries.

After a short debugging phase, we will be able to add new features to the Papillon application:

- Manage private user spaces in order to allow users to annotate (correct) the available entries
- Implement a simple specialized GUI to edit new Papillon entries (as in [9, 10, 11, 12])
- Implement a clean architecture to ease the application localization and provide versions in a lot of languages
- Experiment a WML access to the Papillon database (for cellular phones)
- Integrate tools to check the coherence and ease the management of the Papillon data
- Experiment with XML databases on the data layer.

...

As Papillon is an opened project, this list of extensions will evolve in the time as users will get involved in the specification of the project.

References

- [1] **Blanc E. (1993)** *Visite guidée de PARAX, une base lexicale pentalingue par acceptions sous HyperCard*. GETA, IMAG, 30 p.
- [2] **Mangeot-Lerebours M. (1999)** *Accès unique à des dictionnaires hétérogènes*. Proc. Lexicologie, Terminologie, Traduction (LTT'99), Beyrouth, Liban, 11-13 novembre 1999, AUPELF-UREF, A. Clas ed., 3 p.
- [3] **Mangeot-Lerebours M. (2001)** *Environnements centralisés et distribués pour lexicographes et lexicologues en contexte multilingue*. Thèse, UJF (thèse préparée au GETA, CLIPS), 293 p.
- [4] **Mel'tchuk I. (1981)** *Meaning-Text models: a recent trend in Soviet linguistics*. Annual Review of Anthropology 10 (1981), pp. 27–62.
- [5] **Mel'tchuk I. & Polguère A. (1987)** *A Formal Lexicon in the Meaning-Text Theory: or How to do Lexica with Words*. Computational Linguistics 13/3-4 (1987), pp. 261-275.
- [6] **Mel'tchuk I., Clas A. & Polguère A. (1995)** *Introduction à la lexicologie explicative et combinatoire*. AUPELF-UREF/Duculot, Louvain-la-Neuve, 256 p.
- [7] **Sérasset G. (1994)** *Interlingual Lexical Organisation for Multilingual Lexical Databases*. Proc. 15th International Conference on Computational Linguistics, COLING-94, 5-9 Aug. 1994, 6 p.
- [8] **Sérasset G. (1994)** *SUBLIM, un système universel de bases lexicales multilingues; et NADIA, sa spécialisation aux bases lexicales interlingues par acceptions*. Nouvelle thèse, UJF (Grenoble 1).
- [9] **Sérasset G. (1996)** *Informatisation du Dictionnaire Explicatif et Combinatoire □ le projet NADIA-DEC*. Proc. Lexicomatique et Dictionnaire, Lyon, 28-30 septembre 1996, AUPELF•UREF, A. Clas ed.
- [10] **Sérasset G. & Polguère A. (1997)** *Outils pour lexicographes □ application à la lexicographie explicative et combinatoire*. Proc. RIAO'97, Montréal, 25-27 juin 1997, vol. 2/2, pp. 701-708.

- [11] **Sérasset G. (1997)** *Le projet NADIA-DEC vers un dictionnaire explicatif et combinatoire informatisé* Proc. La mémoire des mots, 5ème journées scientifiques du réseau LTT, Tunis, 25-27 septembre 1997, AUPELF•UREF, A. Clas ed., 7 p.
- [12] **Sérasset G. & Mangeot M. (1998)** *L'édition lexicographique dans un système générique de gestion de bases lexicales multilingues*. Proc. Natural Language Processing and Industrial Applications, Moncton, vol. 1/2, pp. 110-116.
- [13] **Sérasset G. & Mangeot-Lerebours M. (2001)** *Papillon Lexical Database Project: Monolingual Dictionaries & Interlingual Links*. Proc. NLPRS-2001, NII, Tokyo, 27-30 November 2001, pp. 119-125.
- [14] **Tomokiyo M., Mangeot-Lerebours M. & Planas E. (2000)** *Papillon : a Project of Lexical Database for English, French and Japanese, using Interlingual Links*. Proc. Journées des Sciences et Techniques de l'Ambassade de France au Japon, Tokyo, Japon, 13-14 novembre 2000, Ambassade de France au Japon, 3 p.