



Annual Public Report 2010

Table of contents

1. Project description	3
2. Summary of activities	4
3. Project applications	6
4. Dissemination	7
5. Consortium and contact.....	8

1. Project description

FLAVIUS, Foreign LANGUAGE Version of Internet and User-generated Site, is a EC-Project that aims at bridging the language gap between content publishers and users by providing an online platform accessible to websites owners that will enable them to generate multilingual versions of their site, quickly, easily and efficiently in as many languages as they want.

Currently more and more information is published and shared on the internet but is not always available in the working language of each user. Despite the development of translation tools, language is still a barrier, as most people access information written in their own language only.

FLAVIUS project will enable any internet publisher to easily generate multilingual versions of his/her website. The project will produce an easy-to-use platform making an innovative use of automatic translation, human translation, terminology extraction and text correction to cope with the localization of huge amounts of textual information published every day on the internet. It will be as easy to publish in the widespread languages like English, French or Spanish as in less-covered languages like Swedish, Romanian or Polish. On the other side, internet users will be able to better find and understand the information when searching in their own language.

FLAVIUS is thus a step towards creating a web environment where all citizens, companies and governments have an equal opportunity to access information in their own language.



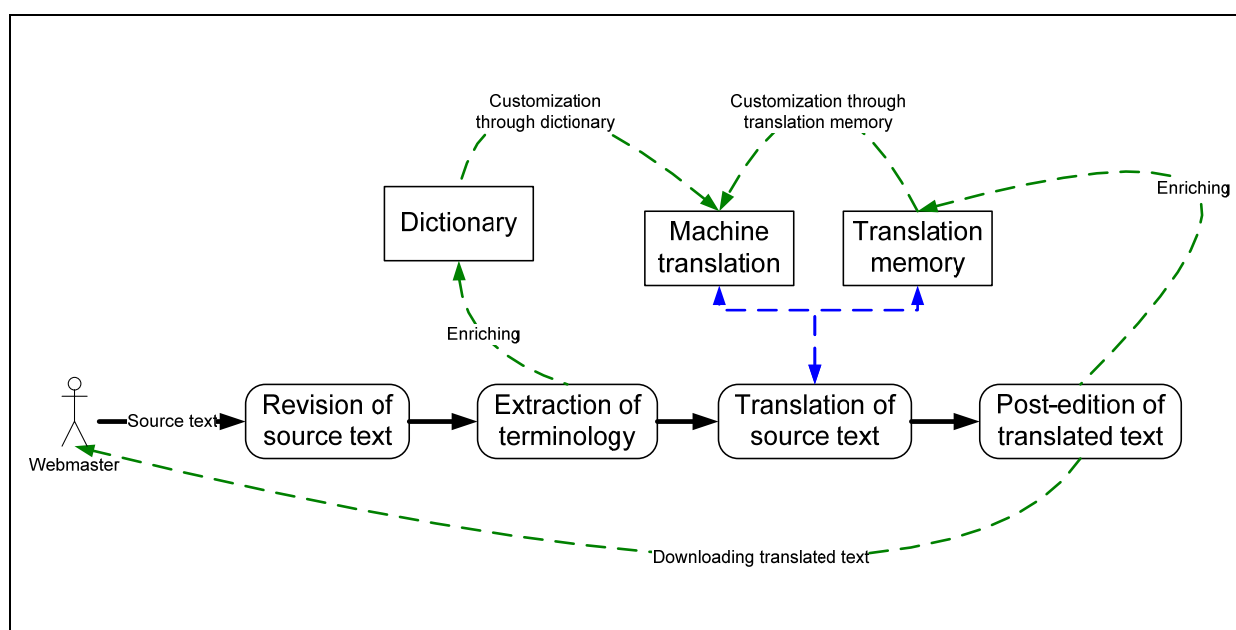
The screenshot shows the Flavius website interface. At the top, there's a header with the project name 'Flavius' and a subtitle 'European project - Foreign Language Versions of Internet and User generated Sites'. Below this is a row of partner logos: Reverso by Softissimo, LANGUAGE WEAVER, DAEDALUS, across, TV trip, QYPE, and overblog. A navigation menu includes 'Home', 'About us', 'Consortium', 'Contact us', and 'Restricted area'. The main content area features a search bar and a featured article titled 'Foreign Language Versions of Internet and User generated Sites', posted on August 5, 2010. The article text describes the project's goal: 'Finding information in one's own language is not an easy task as most websites are not multilingual and remain indexed only in the original language. FLAVIUS aims to provide an easy and cost-effective way for webmasters to have their website translated and indexed into several languages.' A sidebar on the right lists 'Recent Posts', 'Archives', 'Documents', and 'Categories'.

<http://www.project-flavius.eu>

2. Summary of activities

During the first 8 months of the project life cycle, cooperation between the content providers and technical partners has been established, user requirements have been analyzed and the functional features of the FLAVIUS platform have been clearly defined.

Through an in-depth study of existing website translation tools and translation applications, analysis of web technologies and feedback from partners, we gathered relevant information and designed an easy-to-use web platform allowing the translation of websites through the use of optional modules like text correction, translation memory, user dictionaries creation or translation post-edition and able to support a wide range of websites.



Translation of source text through FLAVIUS platform

In relation to this analysis, we defined the platform architecture keeping in mind that one of the key issues for success is the ability to handle a large number of users and a large amount of data without suffering a loss of performance. We also took great attention to design a secure platform to protect user's data.

In order to fulfill these requirements, we have decided to implement in the FLAVIUS platform both the translation of HTML websites and XLIFF files - a standard XML file format in the area of the translation - through a batch processing.

In parallel, we started to adapt the text correction module to the web environment. Written language used by non-expert users in the web scenario poses very different challenges for the correction tools as compared to the "controlled language" that is used by journalists, reviewers or publishers in a professional scenario. Those challenges involve the use of new vocabulary (such as acronyms and abbreviations, slang words, expressions from oral

language) and also new syntax/grammar rules (such as unfinished sentences, different grammar structures, etc.).

As potential users of the FLAVIUS platform are not only professional writers - and can use a lot of acronyms, slang words and expressions from oral language - , text correction will be done in two steps to avoid potential miscorrections. FLAVIUS platform will automatically correct non-ambiguous errors ("My cat are black" will be corrected in "My cat is black"), and suggest corrections for ambiguous errors. The latter will have to be validated through manual correction.

We worked on the customization capabilities of the translation engine as well, as the final quality of the translated texts depends to a large extent on the translation accuracy of the key terms and expressions. It is essential to correctly translate "corporate lawyer" in a business website, and to reuse translations that have already been validated by a webmaster. Translation tools used in FLAVIUS platform will be able to take into account these specificities.

This customization will be done through a dictionary, enriched by the terminology extraction, and a translation memory composed of translations validated and updated by the user of the FLAVIUS platform during the post-edition.

The re-use of already validated translated texts will also solve the issue of incremental translation of content as these translations will be suggested by the Translation Memory each time the webmaster wishes to translate the same content.

Finally, we worked on the aspect relating to the search engine optimization (SEO) as one of the project's goals is to enhance the indexing of websites translated through the FLAVIUS platform.

The terminology extraction tool will be SEO focused and will extract terms with potentially high impact on indexing (HTML title, bold text, etc.). This extraction will give the webmaster the opportunity to suggest the most relevant translation for these terms.



3. Project applications

Website translation and indexing issues

Translation could be seen as "the process of reading, understanding, interpreting, rephrasing and delivering an original message, while capturing all of its subtlety and impact, to a new audience in its mother tongue, in the context of its indigenous culture".

Currently, web contributors write pages in their own language, and publish them directly. In some cases, when they speak a foreign language, they translate some content into this language.

Thus, if the information the user is looking for is not present in the language of the user at the moment he performs his search, it won't be found; try to search for "Red mountains Birmingham Alabama" on Google and then to search for "Montagnes rouges Birmingham Alabama": the fact that the relevant information is not provided reveals that it can only be found with a search in English when the content was written in English. It also demonstrates the need to have translated versions of the current page containing "Montagnes rouges", "Montañas rojas" etc.

The language barrier can only disappear if webmasters make the effort to produce the content themselves. But with 23 official languages in Europe and 22 foreign versions of the website to produce, it is not feasible today to cover all languages. As a result the less-covered languages are always left aside.

Moreover, we couldn't imagine getting all content translated by professional translators for cost reasons. No longer can we use a usual text translator that would translate every piece of text one by one, it would require too much time.

The FLAVIUS platform

All internet content publishers - from webmasters of small websites, who are looking for a large number of visitors and cannot afford to pay for human translation, to large companies, popular websites (including websites with user-generated content), blog platforms, newspapers, ... which intend to reach new foreign customers at a very low cost - could be interested in using the FLAVIUS platform.

In parallel, all internet users will benefit from this by better accessing the online information, even when it has not been written originally in a language they can read.

Translation will not be carried out on-the-fly (as it is the case mostly today) but new multilingual versions of a website will be physically generated by the FLAVIUS platform to be indexed like any other hand-written page by the search engines. Webmasters will just have to provide contents of their site in one language - their native language.

FLAVIUS will combine website architecture management and translation process so that the pages to be published will be automatically organized (with navigation between pages to remain in the same language, etc.) and the webmasters will just have to put them online.

Finally, through customization and collaborative tools, FLAVIUS will enhance the quality of the translation by using webmaster-specific terms and reusing already validated translations.

4. Dissemination

The FLAVIUS project website (<http://www.project-flavius.eu>) is the central place for all potential users to find relevant information about the FLAVIUS platform. Poster, project factsheet and description can be found on the project website.

The following topics are covered:

- Description of the project
- Consortium
- Dissemination documents (poster...)
- Participation in conferences and feedback

FLAVIUS project was also presented at LREC 2010 - an International Conference on Language Resources and Evaluation - in Malta on 19th 20th & 21st May 2010 (EC Projects Village)



FLAVIUS stand at LREC 2010

5. Consortium and contact

The consortium partners have been carefully chosen to bring together the required set of skills.

Source text quality enhancement will be ensured by Daedalus, translation with possible customization will be provided by Language Weaver, collaborative translation memory will be implemented by Across, and the implementation of the FLAVIUS website will be carried out by Softissimo which has a long standing experience in website development.

Another strong aspect of the consortium is the presence of three major European websites (Qype, Overblog and TVTrip) which will provide the “voice of the customer” throughout the project.



Softissimo, FRANCE
<http://www.softissimo.com>



Daedalus, SPAIN
<http://www.daedalus.es>



Language Weaver, ROMANIA
<http://www.languageweaver.com>



Across, GERMANY
<http://www.across.net>



Qype, GERMANY
<http://www.qype.com>



Overblog, FRANCE
<http://www.over-blog.com/>



TVTrip, BELGIUM
<http://www.tvtrip.com/>

Contact:

Christophe Brun-Franc
Softissimo
cbrunfranc@softissimo.com