Partner: Ontologos corp.

ASTECH Project

Statement of actions led
SIXTH FRAMEWORK PROGRAMME
PRIORITY [6]
[Sustainable Energy System]

Contract for:

SPECIFIC SUPPORT ACTION

Project acronym: ASTECH
Project full title: Advanced sustainable energy technologies for cooling and heating applications
Proposal/Contract no.: 19892

WORK PACKAGE 1: TOOLS AND METHODS
# Table of contents

1. **Table of Contents**  

2. **PROJECT SUMMARY**  

   2.1 **ONTologos Corp**  
   2.2 **WP1: Information Technology tools and methods**  
      2.2.1 **Objectives**  
      2.2.2 **Tasks description**  

3. **OSDOC Issues**  

4. **OSDoc Main functionalities**  

   4.1 **Annotation engine**  
      4.1.1 **Supported documents**  
      4.1.2 **Annotation principal**  
   4.2 **OSDoc**  
   4.3 **Search engine**  
      4.3.1 **Request formulation**  
      4.3.2 **Request representation interpretation**  
      4.3.3 **How are the results displayed**  
      4.3.4 **Document details**  
      4.3.5 **MetaSearch engine**  
   4.4 **Knowledge mapping**  
      4.4.1 **Focus a concept**  
      4.4.2 **Improved search**  
      4.4.3 **Control the display**  
      4.4.4 **Display options**  
      4.4.5 **Interpretation of colours**  
      4.4.6 **Access to documents**  
      4.4.7 **Cloud and tree tags**  
   4.5 **Terminology Editor**  
   4.6 **Documents**  
      4.6.1 **Documents posting**  
   4.7 **User account**  
      4.7.1 **Login**  
      4.7.2 **Profile**  

5. **Improvements**  

   5.1 **Terminology improvements**  
   5.2 **OSDoc improvements**  

6. **Conclusion**
2 Project Summary

The objective of this project is to support industrial manufacturers and engineering companies of heating and cooling systems in their development, by introducing more renewable energy sources in their technology. Multi-energy sources, for heating and cooling purposes, are considered to have a large potential in contributing to the penetration of renewable energy sources for domestic, building and industrial applications. But it requires that conventional heating and cooling systems have to be adapted or changed for incorporating renewable energy sources. Nowadays, the heating, ventilating and air-conditioning industries are the actors which have access to the market, and they provide services using mainly fossil fuels or electricity as energy sources. The objective is to have them using renewable energy, combining conventional and renewable sources and proposing environmental cost-effective products.

The technical goal of this project is to set-up technology resource centres in Europe, with the aim of building bridges between the technology providers (research centres, universities and industry) and the technology users (manufacturers, engineering companies…). The project outputs are:

- Creation of the clubs/grouping
- Collection of RTD results
- Development of design tools for evaluating and sizing the new technologies
- Best practice and training programme
- Creation of a knowledge resource centre relying on multi language search engines based on craft ontologies and terminologies
- International collaboration with China

For this 13 partners are collaborating to this project. The consortium is composed of centres of excellence in the area of renewable energies, energy agencies and professional associations representing industry.

2.1 Ontologos corp

Ontologos corp is an Information Technology company dedicated to content management based on enterprise ontology and terminology. A suitable and pertinent information management, e.g. document management, must take into account as well the language in which the information is written as the craft knowledge of the enterprise. It is the reason why the software and the tooled methodology provided by Ontologos corp. rely on linguistic and ontology engineering. All information is classified, or indexed, on the enterprise ontology using a linguistic content analysis. The information retrieval uses multi language search engines which take into account the craft knowledge of the enterprise. The search results are without lost of information and without noise. And last, Ontologos corp. provides graphical interfaces and knowledge visualization for information browsing for users. From the different
solutions proposed by Ontologos corp. let us quote “Os-Doc” a semantic indexing and "intelligent" search engine for information management; “Os-Post” a content analysis engine which forwards all relevant information to the most suitable people (forwarding mails or documents, automatic technological survey, monitoring information…); and “Os-Skill” a knowledge and skill management tool (knowledge sharing, yellow pages, project and experience management…).

The Ontologos corp. technology relies on research results, mainly from the University of Savoie (Condillac Research Group, Pr. Christophe Roche). These results were rewarded by several prices (e.g. the 2001 Rhône-Alps Innovation price with France Telecom) and validated by different industrial applications (EDF Research Centre, VediorBis Foundation, French Education Ministry…).

All Ontologos corp. software applications are provided as web services based on SOAP and XML. They can be directly integrated in end user applications. The terminology building relies on the ISO 704 standard recommendations.


2.2 WP1: Information Technology Tools and methods
Project leader: Ontologos corp.

2.2.1 Objectives

To develop a multi-language search engine, which will be implemented in the websites of the project partners and will be used by the industry first to search as well technical documents as skills or project descriptions, and second to establish collaborations. To inter-connect the partner web-sites and to create a virtual project and technical data-base. Note: As IT tools and methods of the project rely on ontology, let us recall that ontology is a formal description of the concepts of a domain used by a community of practice. The meaning of the words of a terminology relies on the concepts. A shared ontology can be viewed as a common meaning system for multi language terminologies.

2.2.2 Tasks description

- **Subtask 1: Ontology and terminology definition.**

The main goal of this subtask is first to build the common ontology describing the concepts used in RES-HC and HVAC and second to define the multi language vocabularies corresponding to the different partner languages.

- **Subtask 2: Multi language classification and multi language search engine.**
The main objective is to apply an ontology-oriented classification and information retrieval on the multi-language terminologies defined in the subtask1.

- **Subtask 3: Ontology-based graphical interfaces for user browsing information.**

This subtask will study how useful ontology-oriented graphical interfaces are for the users for information browsing.

- **Subtask 4: Website.**

The last subtask will integrate all the previous functionalities into the web site of the project in order to create a common knowledge resource centre.
3 OSDoc Issues

Much more than a passing fad, knowledge management today has become a real challenge for companies.

Aware of the major gaps in classic information retrieval systems, Ontologos corp has developed OS Doc in a very innovative way. The organization and its users are now connected to the process of information management, thereby contributing to its development. The share of your knowledge is now optimal.

Our solution OSDoc can help you get through a single query the most relevant information available within your company.

OSDoc is a search engine based on linguistic and semantic technologies scientifically validated. It incorporates the latest advances in craft terminology and builds on the methodology OUSIA © reference in the field.

OSDoc answers to all of these major problems:

- Organize your knowledge in a structured manner
  - Taking into account your business vocabulary
  - Organizational Approach to apprehend your structure
  - Identifying communities of practices (languages, projects, teams, …)
  - Management of multilingualism
  - Heterogeneous content management
  - Automation of linguistic and semantic analysis of texts
  - Manual indexing of documents possible
  - Dynamic management of your terminology

- Find quickly and easily the relevant information
  - Integration of your job terminology
  - Automatic deepening your queries
  - Innovative interrogation interface
  - Many search parameters and sorting criteria

- Diversify your sources of knowledge
  - Interrogation of your intranet
  - Internet Meta-search engine

- Map and visualise your knowledge
  - Semantic mapping of your terminology
  - Intuitive visualization of your search results
  - Dashboards
4 OSMdoc Main functionalities

Before presenting what we have done with OSMdoc to fit with ASTECH project needs we will recall the main functionalities of this search engine. Remind that OSMdoc is not a simple search engine. It is constructed on lots of semantic and linguistic rules. It is designed to work only with ontologies and theirs terminologies. Its powerful annotation engine allows us to deal with heterogeneous documents and to understand them in an extra-linguistic way.

4.1 Annotation engine

With the ontology, this is probably the most important process of the whole OSMdoc approach. With this process we are able to semantically annotate all your documents. Whatever the language it is written in, we are able to detect all the concepts.

Concepts at the ontology level are totally shared by all the partners. Documents are expressed with terms not concepts. Terms belongs to a language or another. They are semantically shared but not syntactically. The goal of this annotation engine is to abstract documents content to their semantic level. When it’s done, we are able to compare documents not on the words they content but on their informative value.

4.1.1 Supported documents

For now, OSMdoc and its annotation engine are able to deal with:

- Pdf documents (*.pdf)
- Word documents (*.doc, *.rtf)
- Texte documents (*.txt)

4.1.2 Annotation principal

The annotation process can be divided in 3 tasks:

- Extraction of the textual content of documents.
- Lemmatisation of terms. All terms are restored to their definition of the dictionary. We do not have to take into account all the derived forms anymore.
- Similarities search between lemmatized terms and terms of the terminology. We detect all the terms which are equivalent to an item of the terminology. As all terms of the terminology are related to concepts in the ontology, we are able to link documents to ontology by their semantic.

At the end of this annotation we define a structure called “semantic marker”. Each document has its own marker. In this marker we retrieve all the concepts identified. For each concept we know:
• What is its frequency;
• The terms on which we found it;
• Its positions in the document;
• The communities of practice it belongs to (depending on terms on which we found it).
4.2 OSDoc

We will present to you all the main functionalities of the OSDoc search engine. Below a screenshot showing what the search engine interface looks like.

Figure 1: Main interface of OSDoc with default skin
A particular importance has been focused on the ergonomics of the tool. We want to provide sophisticated but accessible software. That’s why the interface is very user-friendly. Independently of the visual skin used, you will always find a navigation bar containing some icons. Each icon corresponds to a specific function.

![Figure 2: Example of navigation bar with the default skin](image)

Here is a list of the functions available:

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Search Engine.</strong></td>
<td>It will look for documents related to your search request.</td>
</tr>
<tr>
<td><strong>MetaSearch Engine.</strong></td>
<td>After a specific work on your request, it will send it to Yahoo! (default web search engine). All the answers will be reannotate with the craft ontology and terminology, re-sorted and then displayed as the standard search engine does.</td>
</tr>
<tr>
<td><strong>Documents.</strong></td>
<td>Access to the documents without formulating any request. They can be sorted according to various criteria.</td>
</tr>
<tr>
<td><strong>Knowledge Mapping.</strong></td>
<td>Here is a visualization of your terminology which allow you to graphically navigate in your concepts network and access documents</td>
</tr>
<tr>
<td><strong>Cloud and tree tags.</strong></td>
<td>Widely present on websites nowadays, this function will show you how your documents are covering your ontology. Some filters allow you to eventually identify experts communities.</td>
</tr>
<tr>
<td><strong>Terminology editor.</strong></td>
<td>Function accessible only if you are logged. Here you are able to update the terminology you are linked to.</td>
</tr>
<tr>
<td><strong>Profile.</strong></td>
<td>Once you are logged, you have access to your personal information. You also can post new documents from here.</td>
</tr>
</tbody>
</table>
You also have a dropdown list always following by a small icon representing flags. This list allows you to change terminology.

Figure 3: List of communities of practice present in the terminology
4.3 Search Engine

The search engine is the main component of OSDoc. That's here you will formulate your search request and then browse the results returned. We will present here several points:

- How formulate a request;
- How interpret your request representation;
- How are the results displayed;

4.3.1 Request formulation

Every search begins with an information needs. Classically formulate through a query in natural language, we tried to provide to users another way to search. Instead of only purpose a blank field and a submit button, we decided to add a tree whose the structure is based on the ontology and where each element is displayed according to terminology of the selected community. Each element (concept) can be expand or collapse if it has children or not (subconcepts). When the user clicked on it, it is automatically added to the current request, which by default is empty. All of this is done with no pages reloading, so it is very easy to see the impact of refining
your search.

The important point is that these two methods can be coupled. You can begin typing words in the blank field, start the search and then add some more concepts with the tree to the current request. After that you are still able to refine the request by the way you preferred.

As you can refine your query, you can also limit it to one or more communities. In fact, by default, the search engine will look after documents related to your request independently of the community they belong to. You can tell the search engine not to keep documents from any given community.

![Figure 6: Choose the community you want to search in](image)

Finally you have some various options. For now, the only one is the possibility to enable or disable the automatic query enrichment. By default, this option is enabled. It will automatically add to your request more specific concepts than them you have selected present in the ontology. If you let this option enabled, you will get more documents related to your request than if you disabled it. The reason is that if you search according to some general concepts, you will probably be interested in documents which are more specific.

![Figure 7: Enable/disable the automatic query enrichment](image)

### 4.3.2 Request representation interpretation

Each request you formulated, whatever the method used, is semantically annotated. As you add concepts to the request, you will see a new part appears located at the top of the area where the results are displayed, just under the blank field. This part presents to you what concepts have been recognized in the request.
We have to prevent you not to be surprised because sometimes it could be a little bit confusing. If you choose to use the blank field to formulate your request, we have to be warned that all terms you wrote matching terms of the terminology will be extracted. After clicking the “GO” button, they will be transferred from the blank field to this details area. Only keywords will remain in the blank field.

Still in this area, you can see that before all concepts of the request you have a small cross circled in red, which allow you to remove the concept from the current search. The trash allows you to fully erase the current request; all keywords and concepts will be removed at one stroke.

4.3.3 How are the results displayed

Lots of efforts have been concentrated on the means to display the documents returned by the search engine. We could not deviate too much from the standards present on the web. Presentations as lists are nowadays widely spread and admitted, users are familiar with such a display.

The way we chose is modern but not disturbing. It was decided to retain this presentation by providing additional visual indicators. You will find:

- A standard description of the document returned (title, summary, uri). You can open the document by clicking on its title.

- In the front of each document, you will see as many colored squares as they are concepts in the request. We vary the opacity of the color to represent the weight of the concept in the document. The balance is done relatively to the whole documents set returned for the search.
• Below the document presentation, you will see a bar which has to be interpreted as the weighted repartition of the concepts of the request inside the document. The bar’s length is normalized in relation to the longest document returned.

• In order to easily grasp long list of results, we provide a duplication of the colored squares on the right of the list. The goal here is to be able to quickly identify in the documents list those which are potentially more interesting for the user. By clicking on the row of squares, we directly go to the document description in the results list.

4.3.4 Document details

Before opening the document, you can open a details window which will present you some information but this time not linked to the request.
In the order, you will find:

- General information as number of words, or the permanent uri...
- A weighted repartition bar which shows you how the concepts extracted during the semantic annotation are disserved in the document.
- A list of the ontological concepts present in the document and their weight.

### 4.3.5 MetaSearch Engine

<table>
<thead>
<tr>
<th>MetaSearch Engine</th>
</tr>
</thead>
<tbody>
<tr>
<td>After a specific work on your request, it will send it to Yahoo! (default web search engine). All the answers will be re annotate with the craft ontology and terminology, re-sorted and then displayed as the standard search engine does.</td>
</tr>
</tbody>
</table>

The metasearch engine included in the OSDoc portal share exactly the same behaviour than the standard search engine. The difference is that the metasearch engine will interrogate indexes of web search engines and not yours. Hence, the semantic annotation work is done on the fly. You have a blank field to fill with your request; once it is done your request is annotated and submitted to a web search engine. In response we get a list of websites (title and summary). We annotate semantically this information to identify which websites are effectively linked to your request and your domain. For now, results are good and we are able to eliminate the noisy results or at less, significantly reduce them. Currently we are able to communicate with Yahoo. On demand, we can configure others search engines.

We do not present the interface because it is the same than the standard engine. The difference here is that you only have a blank field and not all the other options.

Considering the large number of tasks to execute on the fly on the returned results, we can say that the time response of this metasearch engine is pretty good: between 5 and 10 seconds for one request and one web search engine.
4.4 Knowledge Mapping

Knowledge Mapping.
Here is a visualization of your terminology which allow you to graphically navigate in your concepts network and access documents.

Visual tools have always been trendy. Hence we integrated into the OSDoc portal our skills in knowledge mapping. EyeTree is a tool developed by Ontologos corp. It consists on the mapping of your knowledge base. This representation provides a global view of the repository and a very intuitive search. All the nodes are labelled in relation with the selected terminology.

The Eyetree provides several functionalities to control the display:

4.4.1 Focus a concept

Double-click on a node leads to the replacement of this node at the center of the map. Moreover you will only see the hierarchy from this node (its sub concepts) and not the whole one. Another double-click leads back to the normal.
4.4.2 Improved search

This visualization has its own search module. By clicking on “recherche”, a blank field will appear. Just type a few letters inside and the map will automatically move in order to place the occurrence containing these letters at the center.

For example, as a user type the letters “col”, the map moves until the node named “collector” is placed at the center.

4.4.3 Control the display

The “Contrôle” button allows you to move the map.
4.4.4 Display options

- Répartition des éléments...
- Haute Qualité du tracé
- Tronquer les étiquettes
- Centrer

This parameter allows you to choose the way the map will be built: descendant or ascendant.

This parameter allows you to improve the quality of the map. High quality implies higher response time.

This button allows you to restore the default map placement.

This parameter allows you to choose if the labelled have to be truncated or not.

4.4.5 Interpretation of colours

The choice to display some nodes darker than others is not random. The more a node is dark the more there are documents linking the corresponding concept. All documents attached to a concept are simply accessible by right-clicking on the corresponding node. In the submenu “Documents”, you will see a list of documents, select one to open it.

4.4.6 Access to documents

You can access to the documents directly from the EyeTree without formulating any request. Each time a user click on a node, a panel on the right will appears. In this panel the user will have an access to the documents semantically related to the clicked concept. A popup menu available by right-clicking, offers to open the document.
4.4.7 Cloud and tree tags

<table>
<thead>
<tr>
<th>Cloud and tree tags.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Widely present on websites nowadays, this function will show you how your documents are covering your ontology. Some filters allow you to eventually identify communities of experts.</td>
</tr>
</tbody>
</table>

These visualizations are nowadays widely present on the Internet. It consists on a cloud of tags which is supposed to give an idea of what is actually searched on the Internet. It gives a kind of trend.

Also widely biased because of an uncontrolled vocabulary, we decided to build on this representation to show something different. Firstly, the vocabulary is under control. Terms present in these visualizations are taken from the ontology and displayed depending on the selected terminology. Two ideas could be interpreted through this graphs:

- With the cloud, you can see how your documents are covering the ontology. Depending on their font size, you can identify those which are the most important and consequently the most cited by your documents. This way it is possible to see if you have any gaps, just at a glance.

- With the tree, you won’t see quickly the most important concepts of your ontology but the repartition of your documents in the categories of your ontology.

Another very interesting point is the possibility to filter these two displays only under certain communities. You can filter by using the checkboxes on the left. The cloud or the three will automatically resize all terms to take into account your selected

Figure 11: Cloud of tags. Represent the important concepts of the ontology
communities. By this way you will be able to identify which community are the most linked to one concept or to one category. So you can identify experts.

Figure 12: Tree of tags. Represent the repartition of your documents on the categories of the ontology
4.5 Terminology editor

Terminology editor.
Function accessible only if you are logged. Here you are able to update the terminology you are linked to.

OSDoc offers another very useful tool. Recall that all the processes are based on the use of craft ontology and terminology. Users, if they are logged, can by using this tool, update the terminology they belong too. In fact each user is linked to one community of practice, so to one terminology. This software component is very user-friendly. For some homogeneous reasons, you will find the same tree you can use to search documents from the main interface.

![Terminology editor](image)

**Figure 13: Navigation tree to select concepts you want to update**

When you select a node by clicking on it, a form will appear on the right. In this form you will be able to:

- Update the designation of the concept in your terminology. You cannot delete it.
- Add/update/remove synonymous with the intuitive part at the bottom
Each update triggers a backup so all your changes as always saved.
4.6 Documents

Documents.
Access to the documents without formulating any request. They can be sorted according to various criteria.

This part gives a direct access to the documents repository. A simple and standard paged list allows you to open documents. Some sorting criteria are available:

- Sort by name;
- Sort by post date;
- Sort by owner;
- Sort by length;
- Sort by number of references to the terminology.

All documents are displayed as they are in the search engine:

- Its name;
- Its abstract (written during its submission);
- Its URI;
- And other information like its owner, its post date or its last annotation date.

4.6.1 Documents posting

You need to be logged into OSDoc before you can post new documents. See the part dedicated to the user account for more information about the log in task. So once you are logged you have the ability to add some new documents in OSDoc. After their submission, they will be annotated in your terminology. Only after their annotation, they can emerge during users search. Obviously they must have meaning in the terminology.

On your account you will see a part dedicated to the management of the documents you have already posted. An icon describing a kind of hard disk with a green arrow above allows you to add some documents.
By clicking on this small icon you trigger a new post sequence. It consist in several steps.

- Load the document you want to post

![Figure 15: New document loading](image)

Use the “Browse” (parcourir) button to look for your document. Once you found it, use the send file to upload it in OSDoc. When the file has been uploaded, a message will appear saying that the upload is done.

![Figure 16: New document upload done](image)

When you see this message, you can go to the next step by clicking the “Next” button.

- Fill document information

  We need you to complete some general information about the document you are posting. We want you to give us:
  
  - The title of the document in your language;
  - The title of the document in English;
  - The title of the document in your language;
  - The title of the document in English;

  Before to click on the “Finish” button to send the document, do not forget to save it (click on the “save” button).

![Figure 17: Document information form](image)
After clicking on the “finish” button, you will be returned to your account. If all went well, you should see a new document in your list of posted documents.
4.7 User account

User account.
Once you are logged, you have access to your personal information. You also can post new documents from here.

OSDoc has a default behaviour for anonymous users. You can also register some users to give them some advantages in their use of OSDoc. Once a user has been registered by an administrator, he can access to its special functionalities by logging himself in OSDoc.

4.7.1 Login

In the form above, the only thing users have to do is to fill it with their own login and password. If they are registered and their IDs are correct, they will be redirect to their profile administration, otherwise an error message will warn them they are not logged in.

4.7.2 Profile

Figure 18: Login Form

Figure 19: Profile administration page
On this page it will be displayed several pieces of information:

- At the top, you will see some dates and times:
  - The day when your account has been created;
  - Your last connection;
  - The last time your update your account.

- Then, you’ll have general information. It is a simple form so you can update it except your login that is fixed at the account registration.
  - The password is represented with stars to respect the user privacy. If you don’t want to change it, leave it; otherwise don’t forget to confirm it! The password and its confirmation have to be exactly the same. This information can be modified also by an administrator.

- You will also have the possibility, in another part, to post new documents and update/delete those you posted.

You need to be logged in to access to some functionalities:

- Terminology editor;
- Document posting;
5 Improvements

Our work in this project consisted in the adaptation of our tools to essentially fit with the multilingual issue. In fact, at the beginning, Ontologos corp. already proposed a set of tools designed for the conception of craft ontologies and terminologies. Moreover, Ontologos corp. already had a semantic annotation engine and a search engine able to use ontologies and terminologies to manage and reuse domain knowledge.

The task related to the ASTECH ontology conception has been documented in several reports written by Ontologos corp. during the whole project duration. Hence, we will only present here the actions we led to adapt our tools for this specific configuration:

- The terminology environment had to be adapted to take into account the multilingualism aspect and by the way to allow an extra-linguistic means of communication. We provided a simple web version of this editor which has been integrated in the search engine. This way, users can dynamically enrich and/or update their terminology.
- The main interest of the semantic annotation tool is to detect not only the keywords but the concepts. The existing engine must now consider the multilingual.
- The existing search engine had to be improved in several ways. First, it had to integrate the multilingualism terminology in its processes. Then, the interrogation interface should be multilingual. You have to understand that it is not only a translation we did. Some items are displayed depending on the craft terminology. We also had to graphically adapt our search engine to match with the ASTECH visual chart. Remind that our search engine had to be integrated in the ASTECH web portal.

5.1 Terminology improvements

Before starting this project, we used our tool named TCW\(^1\) to develop the craft terminology. It consists on a kind of language layer over the ontology. Each item of the terminology is linked to a concept of the ontology. This item is actually a set of synonymous which all designed the ontological concept.

In the case of the ASTECH project we had to update this layer introducing new elements. We are face to a consortium of 13 partners which all share a common knowledge base: the ontology. Remind that the main issue here is to allow the sharing of craft knowledge among these partners. We know that ontology is not extra-linguistic, so we introduce the concept of community of practice to isolate each

\(^1\) Terminology Craft Workbench
partner and its own terminology. In fact, each partner, beyond its language, have its own words and synonymous. At this stage, we have a consensual ontology and as many terminologies as partners.

The fact that all terminologies items are linked to concepts located in the shared ontology make possible the dialog between the different communities. We are this way able to say which synonymous are equivalent independently of the language they are written in.

5.2 **OSDoc improvements**

The search engine we have adapted is named OSDoc. We'll see later in this document what are its main functionalities. For now, we just want to present what we have updated.

As we say before, its main functions like the semantic annotation engine or the search engine have evolved towards mechanisms able to deal with several communities of practices (terminology). Semantic annotations are still driven on the terminology but resulting conceptual metadatas associate in addition the communities identified.

The figure below illustrates the principal of annotation. All documents are assigned a semantic marker where each founded concept is linked to a community of practice.

With such improvements, it is now possible to formulate a search request in English and get in response some documents written in another language but semantically linked to the information needs.

The search engine gives access to a lot of parameters and filters which allow the user to give some limits to its request. For example, he can:
Choose the communities he want to include in its search;
Choose the kind of documents returned by the search engine;
Choose the way its request will be treat by the search engine;
...

There are on the OSDoc interface lots of elements displayed which could be simple labels or references to the terminology. These two aspects are important because with OSDoc we have two notions of language: the spoken one and the terminological one. An English person can display all the labels in English but he can change the terminology. In the particular case of this project where the communities of practices have been identified depending on their spoken language, changing the language or the terminology is the same.
6 Conclusion

Today, OSDoc is fully functional and accessible from the ASTECH website. Here is a list of some characteristics figures:

- The ontology built integrates more than XX concepts;
- The terminology has 13 communities of practices;
- There are more than XX synonyms in this terminology spread through the communities.
- At this time, we manage about XX documents;
- ...