

# **With WORLDTREK family, create, update and browse your terminological world**

**Yasmina ABBAS\*, Marie-Luce PICARD\*\***

\*EDF – Division Recherche & Développement (Division R&D)  
MTI/NTIC/TAIC

01, Avenue du Général de Gaulle  
92141 CLAMART Cedex, FRANCE

INALCO – Centre de Recherche en Ingénierie Multilingue (CRIM)

01, Rue de Lille  
75007 PARIS, FRANCE

yasmina.abbas@edf.fr

\*\*EDF - Division Recherche & Développement (Division R&D)

SDC/CLEO/SOAD

01, Avenue du Général de Gaulle  
92141 CLAMART Cedex, FRANCE

marie-luce.picard@edf.fr

## **Abstract**

Companies need to extract pertinent and coherent information from large collections of documents to be competitive and efficient. Structured terminologies are essential for a better drafting, translation or understanding of technical communication. WORLDTREK EDITION is a tool created to help the terminologist elaborate, browse and update structured terminologies in a ergonomic environment without changing his or her working method. This application can be entirely adapted to the « terminological habits » of the expert. Thus, the data loaded in the software is meta-data. Links, status, property names and domains can be customized. Moreover, the validation stage is facilitated by the use of templates, queries and filters. New terms and links can be easily created to enrich the domains and points of view. Properties like definition, context, equivalent in foreign languages are associated with the terms. WORLDTREK EDITION facilitates the comparison and merging of pre-existing networks. All these tasks and the visualization techniques constitute the tool which will help the terminologist to be more effective and productive.

## **1 Context**

The increasing use of the internet provides companies with constantly growing collections of information. These Megabytes of electronic documents are added to the knowledge capital of organisations (plans, technical manuals, models, reports, papers gathered in databases).

The ultimate purpose of collecting this overwhelming amount of information is competitiveness. To serve customers well and be efficient, companies must shorten their product development time, innovate and deliver high quality products by exploring their experience and the experience of others. None of this is possible without retrieving accurate and coherent information from the documents collected.

In this particular case, structured terminologies are essential. They can be used for a better understanding,

drafting or translation of documents. A network of terms enriched with properties like context, definition and synonyms constitute a significant help when combined with tools like electronic books or computer-aided translation systems.

## **2 Terminology production in EDF**

The largest energy supplier in France, Electricité De France (EDF), conceives, exploits and maintains power plants. In the field of textual-document managing, the company faces difficulties. On the one hand, the collections of technical communication describing these power plants need to be coherent with each other (ABBAS et al., 1999). On the second hand, the opening of the electricity market in Europe increased EDF needs in multilingual translation and business intelligence.

One of the activities of the Information and Knowledge Advanced Processing group (Traitements Avancés de l'Information et des Connaissances) in the Research and Development Division of EDF is to implement tools helping the expert consult and produce technical documents (BOCCON-GIBOD, 1999). One of them, WORLDTREK EDITION, enables to create, update and browse terminologies.

### 3 Purposes

WORLDTREK EDITION (PICARD et al., 1999) is an application intended for experts or terminologists. It was conceived to help them validate and enrich a terminology through the use of graphical visualization techniques.

Lists and graphics are used to create terms and links, structure the terminology by adding domains and points of view. The validation stage is also facilitated by the use of templates and queries.

#### 3.1 The users

A terminology can be defined as a structured set of designations and concepts in a specific field. This network of specialised terms linked together and enriched with properties can be elaborated following different approaches and theories. The terminologist organizes his or her work taking into account his or her theoretical background and his or her experience.

WORLDTREK EDITION was conceived to assist the terminologist. One of our main objectives is the customization of the tool according to the user's needs. All the software functionalities were set to follow the different stages of terminology construction. WORLDTREK EDITION aims at being an ergonomic tool which may improve the expert's productivity.

The expert uses this application to organize his or her work according to the results he or she expects. Defining concepts, validating terms, creating domains, points of view and links depend on the choices of the terminologist. WORLDTREK EDITION is generic enough to be applied for describing and exploring a technical field, building a representation, enriching, updating and comparing it with other terminology models.

#### 3.2 The Input data

Before presenting a description of WORLDTREK EDITION and its various functionalities, a presentation of the input data is essential.

WORLDTREK EDITION exploits the results of LEXTER (BOURIGAULT, 1994), a terminology retrieval software. The basic principle is splitting the text by locating terminological noun phrases boundaries. It performs a morpho-syntactical analysis of a corpus of French texts on any technical field.

The candidate terms retrieval processing is composed mainly of two steps : a splitting one and a parsing one. The splitting module is in charge of splitting the text by locating potential boundaries between which noun phrases likely to be occurrences of terms are isolated. Then, at the

parsing stage, LEXTER parses the maximal-length noun phrases in order to grammatically break up each complex candidate term into a head and an expansion. Each candidate term is associated to morphological information: grammatical category, morphological features (particularly gender and number) and lemmatized form.

However, WORLDTREK EDITION can exploit any output given by any other terminology retrieval software.

### 4 WORLDTREK EDITION principles

WORLDTREK EDITION was elaborated to help the expert or the terminologist build up a structured terminology rapidly and effectively. The various functionalities of WORLDTREK EDITION are adapted to the following tasks : modelisation, validation, structuring, comparison of networks and data export. These steps are presented through an overall description of the software. In the next section, an example of terminology elaboration gives a concrete view of WORLDTREK EDITION tasks.

#### 4.1 Modeling

The data model loaded in WORLDTREK EDITION is a meta-model. The terminology is composed of a set of elements which can be modified.

The terminology model can be totally created by the expert. The terminologist defines and creates domains, points of view, types of links (synonymy, hyponymy, hyperonymy), properties like definition, grammatical category and status labels designating term validity.

The terminologist can also reuse a pre-existing model and adapt it to a new data model.

#### 4.2 Validating

The validation stage concerns two items in a terminology elaboration, terms and links. The expert has to choose the terms which will be closer to the conceptual representation of a technical field. He or she will also select the types of links which will relate terms to each others.

The expert has also the possibility of assigning a status to terms and links. This functionality allows to give additional information on the nature of a term or a link. An expert in charge of drafting technical communication will be able to choose between two apparently similar terms by consulting their status. For example, the control room which controls nuclear power plants is « salle de conduite de tranche » for the Paluel series and « salle de contrôle-commande » for the N4 series, a new generation of nuclear facilities. The status permits the technical writer to select the right designation according to the documents he or she is drafting by indicating which term is more appropriate. The use of « salle de conduite de tranche » will be forbidden when drafting technical communication concerning The N4 series. The status of terms can be defined and adapted to the final purposes of the terminology.

WORLDTREK EDITION facilitates the validation of a terminology by several experts with the customization of

the term status. A power-plant terminology is connected with many different technical fields, electricity, electronics, mechanics or physics. Status enables to follow the validation of terms by different specialists.

WORLDTREK EDITION provides different means of validating terms and links: templates, a modification form or the contextual menu of terms and links. Moreover, queries and filters help select terms according to their status.

### 4.3 Structuring

The structuring of a multidisciplinary terminology is not an easy task. With WORLDTREK EDITION, the expert can create different domains and assign them to several points of view. This functionality allows to elaborate a multi-faceted terminology. For instance, a terminology describing a nuclear power plant would contain domains describing its different elementary systems or buildings. On the other hand, the points of view would reflect the maintenance, safety and the construction aspects.

The structuring stage involves the creation of new terms and links to enrich the different domains. Templates can be used to facilitate this task. It allows the user to propagate the same criteria to a set of terms or links. These functionalities can be used to validate terms in a list or in a graphic and to assign them to the domains of the terminology.

One point of view can gather one or several domains and one domain can be assigned to many points of view.

### 4.4 Enriching

A terminology is not just a network of terms organised in groups of domains and points of view. It can be enriched with properties. This varies according to the final use of the terminology.

Grammatical category, lemma, frequency in the text constitute the most common properties. Nevertheless, other elements can be added. A terminology which will be used in a computer-aided translation system will contain the term equivalents in one or several foreign languages. In an application conceived to assist the drafting of technical communication, the definition and the developed form of abbreviations will help the technical writer to choose the proper term.

### 4.5 Merging networks

WORLDTREK EDITION is not only used to create a new terminology but also to update it. The expert is able to add new information and change the structure of the network as explained in the modelling section.

The new information can also be extracted from an existing network of terms. A validated terminology is imported and the terms are added to the list. The terminologist is able to compare terms for example to check the labels and harmonize them. This task is then followed by the retrieval of doubles to clean up the network.

With the merging functionality, the user applies the properties, the links and the status of terms to other ones in the network.

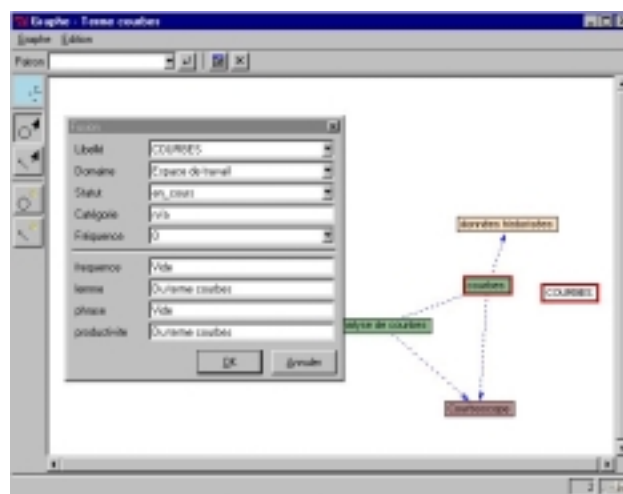


Figure 1. Term merging

The graphic facilitates the comparison between networks and highlights similar junctions.

Thus, different terminologies already validated and structured by several experts can be merged and exploited, which shortens effectively the production and maintenance time.

### 4.6 Exporting

Export is achieved by using retrieval options. The user can extract a group of terms according to some criteria like the domain they belong to, the closest terms they are related to, their status or their properties. Lists are constructed and then exported in XML format.

At this stage, term retrieval can also be carried out in the graphic. Small groups of terms related to each others can be selected, then converted into lists in order to be exported. The whole structured terminology can also be exported.

## 5 WORLDTREK EDITION functionalities

WORLDTREK EDITION functionalities are adapted to the different stages in terminology elaboration. A concrete example shows how the different functionalities help the expert to validate, structure, enrich and export terms collected in a corpus of technical texts. The network of candidate terms loaded in WORLDTREK EDITION for the example was retrieved from a document which presents the activities, the members, the partnerships and the projects of the SOAD group (Statistics, Optimization and Data Analysis) at the Research and Development Division of EDF.

The example below presents the functionalities of WORLDTREK EDITION in a certain chronology, which can be changed according to the expert's methodology.

WORLDTREK EDITION displays the network of terms in a list or a graphic. The interactions between the

two enable the expert to construct the terminology effectively.

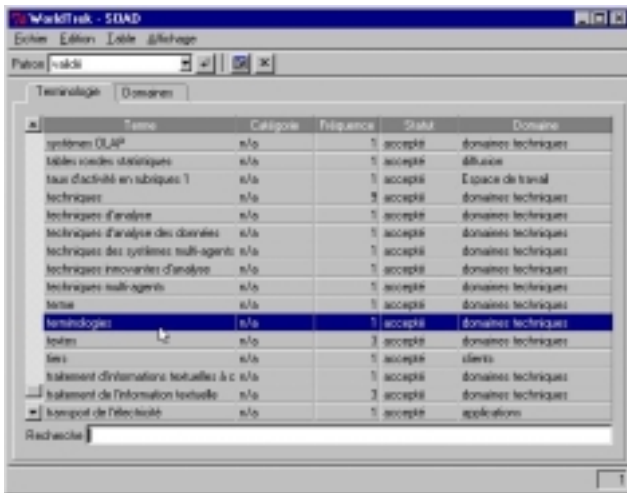


Figure 2. Term list

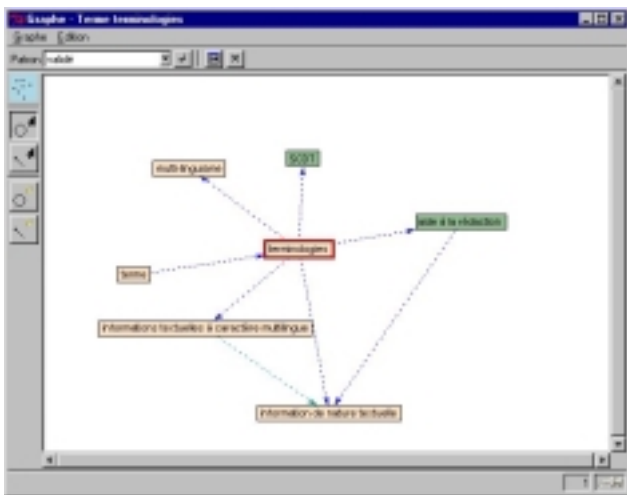


Figure 3. Graph

### 5.1 Creating the model

The expert defines the model of the future structured terminology. In this example, the user captures the different status labels in a model form to be able to select them when constructing a template, a query or consulting the contextual menu.

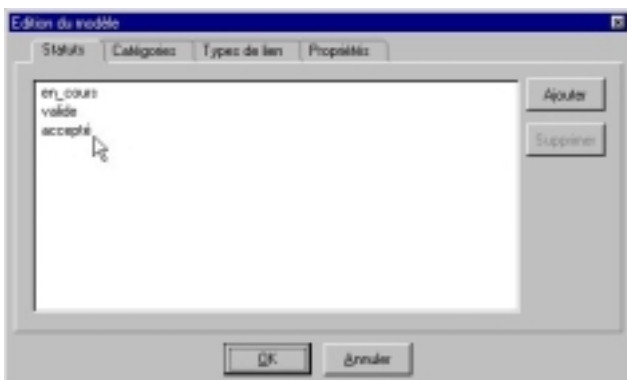


Figure 4. Terminology model

Property and grammatical category labels are added when enriching the terminology. The expert captures link labels as he or she creates them during the structuring stage. In this terminology, status may have one of the three values : « yes », « no » and « to be seen ».

### 5.2 Loading and validating candidate terms

The candidate terms are retrieved with LEXTER, collected, then loaded in WORDLTREK EDITION. At this stage, the user has to validate and select the terms of the future structured terminology.

The status « not validated » is the default status. The user can validate the candidate terms in two ways :

- One by one in the list,
- As a group of candidate terms, by selecting them with the mouse and using a template to assign them a value,
- As a group of candidate terms by using a modification form.

Templates are easily constructed and not only used to validate terms. They can be adapted to different tasks : the creation of new terms and links, the modification of status and properties or the assignment of terms to domains.

The user defines the template's criteria according to the current task. In this case, he or she assigns a validation status to each term. Different templates are constructed, one for the « yes » status, another one for the « no » status and a last one for the « to be seen » status.

The modification form also enables the validation of a group of candidate terms. The modification form is temporary and can not be saved as it is the case of templates. Clicking on a button in the list opens it then the user selects the different modification criteria.

After the validation, queries or pre-existing filters on the status are used to check if all the candidate terms are validated.

In the graphic, the expert customizes his or her terms by attributing different fonts according to the status. In this way, he or she can detect more easily the validated terms from the candidate terms.

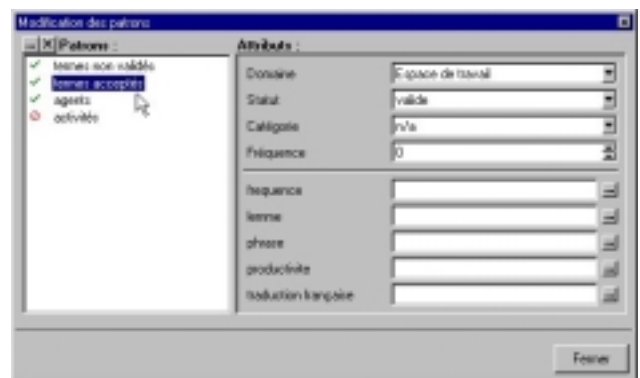


Figure 5. Template

### 5.3 Creating domains and points of view

The creation of domains and points of view are part of the structuring stage. The terms extracted from the

document must illustrate the different entities of the group and its activities. The domains chosen reflect the structure of the document analysed and processed : group members, projects, technical fields, partnerships, clients, etc.

The user creates a new domain by attributing a name, a number, a colour and one or several points of view. This customization will help him navigate in the graphical network, easily detect the domains and control term distribution. The points of view are used especially in large networks of terms to facilitate de navigation in a wide-ranging and multidisciplinary terminology.

#### 5.4 Assigning terms to domains

After creating the domains, the expert assigns the validated terms and relates them to each others with different types of links. New terms can be added to complete or enrich the current domains.

The terminologist can use the same methods presented in the validation section :

- Templates,
- Modification form,
- One-by-one assignment with the contextual menu.

There are slight differences between the three methods. As it is the case for validation, templates can be elaborated and saved. They can be reused, unlike the modification form, which is temporary. The user can also modify terms individually by using their contextual menu.

Templates are also used to create new links between terms. In this example, the terms are already related to each other with head and expansion links created by LEXTER. For instance, a term like « reactor coolant system » is decomposed in a head « reactor » and a expansion « coolant system ». However, the user can insert other links like synonymy, hyponymy and hyperonymy. He or she selects his or her terms in the list or in the graphic and propagates the template.

The expert customizes the various links with different colours and shapes to distinguish them in the network.

Queries and filters enable the user to extract doubles and check if terms are not isolated in the network.

In WORLDTREK EDITION, terms morphologically similar or syntactically close are detected and a list is created. The user analyses the terms and deletes or merges items according to their similarity, their links, ... When two terms are merged, the links of the first one are assigned to second one. These functionalities are used to check if the network is coherent.

#### 5.5 Assigning properties

The network created by LEXTER, validated and structured by the user already contains properties. Grammatical category, lemma and frequency in the corpus are displayed in the contextual menu of each term. However, the user can easily and freely add properties. He or she must add the new properties' labels in the terminology model to declare them.

The new properties are entered for each term. The templates, the modification form or the contextual menu

are used to associate the information with the corresponding terms.

#### 5.6 Exporting data

The structured terminology is exported in XML format.

The user creates lists of terms by using queries or filters and exports them according to his or her criteria. Domains can also be exported individually. Moreover, the expert can choose the properties.

### 6 WORLDTREK EDITION perspectives

The evolution of WORLDTREK EDITION is closely related to the final applications of structured terminologies.

Terminologies are also used to create indexes and facilitate full text research in electronic books. WORLDTREK EDITION is about to be updated in order to display the original text from which terms are retrieved. This connection to the corpus facilitates validation. The expert needs to consult the context in order to assign a status to the term.

WORLDTREK EDITION is used within data mining applications. It can help the user to explore his or her databases and will be useful in constraint checking.

Structured terminologies created by WORLDTREK EDITION can be used by WORLDTREK, a tool which offers interactive graphical display of the semantic network and possibilities of graphical retrieval and browsing. This application is part of authoring and comprehension aid systems. With WORLDTREK, the user can easily find a term, its definition or control its validity. This tool is mainly exploited to consult and browse terminologies.

### 7 Acknowledgements

We would like to thank Mrs Daniela Garcia, Mr Eric Boudaillier and Mr Didier Bourigault for their precious remarks.

### 8 Bibliographical references

ABBAS, Y. & PICARD, M.-L. (1999). *Exemple de pratique terminographique en entreprise*. TIA 99', Nantes, France.

BOCCON-GIBOD, H. (1999). *Enjeux de la maîtrise de la terminologie pour la production et la consultation de documents – Application à la documentation technique des installations de production d'électricité*. DocuWorld 1999, Paris.

BOURIGAULT, D. (1994). *LEXTER, Un Logiciel d'Extraction de TERminologies-Application à l'acquisition des connaissances à partir de textes*. Thèse de l'École des Hautes Etudes en Sciences Sociales, Paris.

PICARD, M.-L. & BOUDAILLIER, E. (1999). *Worldtrek for authoring and comprehension*. IUI'99, Redondo Beach, Californie, USA.