

Aporte del dominio de la terminología a la producción y consulta de documentos

Daniela GARCIA, Daniela.garcia@edf.fr
Marie-Luce PICARD, ml.picard@der.edf.fr

EDF Division Recherche et Développement, 1 Avenue du Général de Gaulle 92 140 CLAMART - Francia

El trabajo presentado es llevado a cabo por un grupo compuesto por H. Boccon-Gibod, M.-L. Picard y D. García, ingenieros-investigadores de la R&D de EDF, a cargo de la definición y la implantación de herramientas para aplicaciones de proceso de información textual. Didier Bourigault es investigador del CNRS (ERSS-Toulouse) y coopera con la Division R&D en la extracción y estructuración de terminologías.

1. Contexto y problemática

Las instalaciones de producción de electricidad, en particular las instalaciones nucleares, tienen un ciclo de vida largo y son descritas en documentos muy extensos [1]. Tienen una gran exigencia de seguridad; eso implica la necesidad de asociarles documentos técnicos de muy alta calidad.

Intervienen numerosos actores, y hay también numerosas fuentes de información: planos, esquemas, textos. Desde los años 70 hasta nuestros días, las tecnologías han evolucionado: la implantación de "sistemas de información" ha sustituido a los planos trazados a mano. Sin embargo, la mayor parte de la información resta bajo forma de textos redactados en lenguaje natural, asociados a esquemas (en dos dimensiones) y a modelos (en tres dimensiones) almacenados en bases de datos. De esta manera, la problemática de producción y de consulta de esos documentos técnicos permanece enteramente, con una dimensión multilingüe, puesto que EDF vende actualmente su experiencia y sus servicios en mercados internacionales de ingeniería. Debemos acentuar nuestro dominio de la producción interna de documentos con el fin de asegurar: la **coherencia** al interior de un mismo documento, entre varios documentos, entre documentos y datos de referencia (modelos, esquemas, planos); la **calidad**; la velocidad de escritura; las traducciones. Debemos también

proponer accesos eficaces e inteligentes al **contenido** de un documento. En fin, debemos integrar el puesto del autor y el sistema de consulta en los sistemas de informaciones técnicas. Para producir documentos de buena calidad, en varias lenguas, comprenderlos mejor y más rápidamente, y finalmente para ganar tiempo en la búsqueda de informaciones, es esencial una **terminología**.

Esta problemática es general y se aplica a todos los sectores de la actividad: las preocupaciones por la productividad, la eficacia, y la reducción de tiempos de lanzamiento al mercado.

2. El proceso y las herramientas

2.1 Terminologías estructuradas

Una terminología estructurada permite unificar la representación del contenido textual de los documentos con la representación formal de los modelos subyacentes. De esta manera los términos técnicos utilizados deben corresponderse de manera precisa con los modelos de datos: **coherencia** asegurada! Los modelos de datos proporcionan igualmente entradas (índices) para la consulta de documentos técnicos: **eficacia** asegurada!

2.2 Elaboración automática de terminologías

Disponemos de varias herramientas para la constitución automatizada de terminologías estructuradas :

_ **LEXTER** [3 y 3bis] procesa textos y extrae términos candidatos,

_ **FASTR** [4] propone familias de sinónimos,

_ Una interfaz de validación (**HTL**, sobre **ACCESS**) permite seleccionar a mano los términos pertinentes y asociarlos entre ellos.

_ **WORLD TREK** [2] permite visualizar la red de términos, explorarla gráficamente, y pronto modificarla y validarla gráficamente.

La **estructura** de la terminología es esencial, y está fuertemente conectada a su utilización posterior y al público al que está dirigida [5]. Los vínculos necesarios son *de hiponimia/hiperonimia* y *de sinonimia*. Las herramientas surgidas de trabajos de tesis nos proporcionan preciosas ayudas para la asistencia a la estructuración: COATIS [6 y 6bis] proponiendo vínculos de causa-efecto entre “términos de acción”; LEXICLASS [7 y 7bis] vía la constitución de clases de términos con una semántica homogénea. La atribución de statuts a los términos es también fundamental (invalidado, validado y autorizado, validado pero prohibido...) para una explotación eficaz de la red terminológica.

2.3 Puesto del autor

La integración de la terminología en el puesto del autor permite **controlar en la fuente la calidad de los documentos** :

_ garantía de una terminología coherente (sobre todo cuando hay múltiples autores), y en sus traducciones,

_ optimización de la calidad (precisión / coherencia),

_ aumento de productividad (evitar las modificaciones del texto a posteriori, facilitar los procesos de control).

En cooperación con XEROX, proponemos un prototipo de puesto de ayuda a la redacción, **Merlin** : un corrector terminológico (*Xerox TermChecker*) integrado al tratamiento de textos (como un corrector ortográfico) y asociado a WORLD TREK para la navegación gráfica en la red terminológica. Redactores y lectores verifican inmediatamente (y en su entorno de trabajo habitual) el empleo correcto de un término técnico en un contexto preciso, tienen acceso a la traducción oficial de un término, y disponen fácilmente de la definición de un término.

3. Conclusión

Las terminologías estructuradas (y multilingües) son un aporte mayor al dominio de la documentación de empresas. Llegando al año 2000, disponemos de un proceso y de herramientas para constituir las y explotárselas eficazmente. Sin embargo, la función del (de la) terminólogo(a) sigue vigente : para utilizar bien estas herramientas, ayudar a los expertos en su proceso de validación, actualizar informaciones terminológicas.

Agradecemos a Yasmina Abbas, ingeniera terminóloga que trabaja para EDF-Division R&D, a Beate Lippold, François Cachard, Jean-Marc Coursimault y Cyril Chantrier de la sociedad XEROX, a Eric Boudaillier de la sociedad UNIFIX por su trabajo sobre WORLD TREK, y a Gustavo Crispino por ayudarnos a traducir este resumen.

Bibliografía

[1] H. Boccon-Gibod. « Enjeux des outils terminologiques pour la conception, la production et la consultation des documents – Application à la documentation technique des installations de production d’électricité », *DocuWorld 1999*, Paris, 1999.

[2] M.-L. Picard, E. Boudaillier. “Multilingual WorldTrek for authoring and comprehension”, *IUI’99*, Redondo Beach CA, 1999.

[3] D. Bourigault. *LEXTER, un Logiciel d’EXtraction de TERminologie. Application à l’acquisition de connaissances à partir de textes*. Thèse EHESS, Paris, 1994.

[3bis] D. Bourigault. “LEXTER, a Natural Language Processing Tool for Terminology Extraction”. In *Proceedings of the 7th EURALEX*. Göteborg, 1996.

[4] C. Jacquemin. *Variation terminologique : reconnaissance et acquisition de termes et de leurs variantes en corpus*. Habilitation. Université de Nantes, 1997.

[5] Y. Abbas. *La doctrine terminologique et la pratique terminographique*, Mémoire de DESS. INALCO/CRIM, 1999.

[6] D. Garcia. *Analyse automatique de textes pour l’organisation causale des actions, réalisation du système informatique COATIS*. Thèse de l’Université Paris 4, 1998.

[6bis] D. Garcia. "COATIS, an NLP system to locate expressions of actions connected by causality links". In *Proceedings of the 10th EKAW*. In Springer's Lecture Notes in Artificial Intelligence. Sant-Feliu-de-Guixols, Catalogne, 15-18 octobre, 1997.

[7] H. Assadi, *Construction d'ontologies à partir de textes techniques ; application aux systèmes documentaires*. Thèse Université Paris 6, Paris, 1998.

[7bis] Cécile Gros, Housseem Assadi. "Technical Documentation Consulting Hypertexts : General Architecture and an Example of Implementation". In *Proceedings of Hypertext'98*, Pittsburgh, USA.