



TOTh 09

Terminologie & Ontologie : Théories et Applications

Actes de la troisième conférence TOTh - Annecy - 4 & 5 juin 2009



Institut Porphyre
Savoir et Connaissance

Dans un monde où la communication et le partage d'information sont au cœur de nos activités, les besoins en terminologie se font de plus en plus pressants. Il est devenu impératif d'identifier les termes employés et de les définir de façon consensuelle et cohérente tout en préservant la diversité langagière.

La terminologie, en tant que discipline scientifique, se fonde sur une conceptualisation d'un domaine et sur les mots pour en parler. Elle se doit donc de concilier un point de vue linguistique et un point de vue ontologique. Elle doit également, dans une société numérique où les connaissances constituent la principale richesse, pouvoir être opérationnalisée à des fins de traitement de l'information.

Les conférences TOTh se situent dans le prolongement des colloques annuels de la Société française de terminologie organisés en décembre à Paris (Ecole normale supérieure de la rue d'Ulm). Planifiées à mi-parcours, au mois de juin à Annecy (Polytech'Savoie), elles en complètent l'offre et proposent des conférences avec appel à communications, comité de lecture et publication des actes.

Les conférences TOTh ont pour objectif de rassembler industriels, chercheurs, utilisateurs et formateurs dont les préoccupations relèvent à la fois de la terminologie et de l'ontologie et, de façon plus générale, de la langue et de l'ingénierie des connaissances. Elles se veulent un lieu d'échange et de partage où sont exposés problèmes, solutions et retours d'expériences tant sur le plan théorique qu'applicatif ; ainsi que les nouvelles tendances et perspectives des disciplines associées : terminologie, langues de spécialité, linguistique, intelligence artificielle, systèmes d'information, ingénierie collaborative, etc.

Christophe Roche, Président du Comité Scientifique

<http://www.porphyre.org>



Institut Porphyre
Savoir et Connaissance

ISBN 978-2-9536168-0-4
EAN 9782953616804

Publications précédentes

TOTh 2007

Actes de la première conférence TOTh - Annecy - 1^{er} juin 2007

TOTh 2008

Actes de la deuxième conférence TOTh - Annecy - 5 et 6 juin 2008

Commandes à adresser à : toth@porphyre.org

Titre : TOTh 2009. *Actes de la troisième conférence TOTh - Annecy - 4 & 5 juin 2009*

Editeur : Institut Porphyre, *Savoir et Connaissance*

<http://www.porphyre.org>

Annecy, 2009

ISBN 978-2-9536168-0-4

EAN 9782953616804

© Institut Porphyre, *Savoir et Connaissance*



Actes de la conférence

TOTh 2009

Annecy – 4 & 5 juin 2009

avec le soutien de :

- Société française de terminologie
- Association Européenne de Terminologie
- Ecole d'ingénieurs Polytech'Savoie – Université de Savoie
- Université de Sorbonne nouvelle
- Association EGC (Extraction et Gestion des Connaissances)
- ISKO (International Society for Knowledge Organization) France



Institut Porphyre
Savoir et Connaissance

<http://www.porphyre.org>

Comité scientifique

Président du Comité Scientifique : Christophe Roche

Comité de pilotage

Loïc Depecker	Professeur, Université de Sorbonne nouvelle
André Manificat	Directeur, GRETh
Christophe Roche	Professeur, Université de Savoie
Philippe Thoiron	Professeur émérite, Université de Lyon II

Comité de programme

Bruno de Bessé	Professeur, Université de Genève
Pierre Blanc	EDF SEPTEN
Danièle Bourcier	CNRS, CERSA Paris
Marc van Campenhoudt	Professeur, Termisti, ISTI, Bruxelles
Danielle Candèl	CNRS, Université Paris Diderot
Stéphane Chaudiron	Professeur, Université de Lille III
Viviane Cohen	France Télécom, Paris
Rute Costa	Professeur, Université Nouvelle de Lisbonne
Luc Damas	MCF, Université de Savoie
Sylvie Desprès	MCF, Université Paris XIII
François Gaudin	Professeur, Université de Rouen
Anne-Marie Gendron	Chancellerie fédérale suisse, Section de terminologie
Jean-Yves Gresser	ancien Directeur à la Banque de France
Ollivier Haemmerlé	Professeur, Université de Toulouse
Jean-Paul Haton	Professeur, Université de Nancy 1
Michèle Hudon	Professeur, Université de Montréal
John Humbley	Professeur, Université Paris 7
Michel Ida	Directeur MINATEC, CEA
Hendrik Kockaert	Professeur, Lessius Hogeschool (Anvers)
Michel Léonard	Professeur, Université de Genève
Pierre Lerat	Professeur honoraire, Université Paris XIII
Widad Mustafa	Professeur, Université de Lille III
Henrik Nilsson	Terminologocentrum TNC, Suède
Jean Quirion	Professeur, Université du Québec en Outaouais
Renato Reinau	Suva, Lucerne
François Rousselot	MCF, Université de Strasbourg
Gérard Sabah	CNRS, Orsay
Michel Simonet	CNRS Grenoble
Marcus Spies	Professeur, Université de Munich
Dardo de Vecchi	Professeur associé, Euromed-Management

Comité d'organisation :

Responsable : Luc Damas
Samia Chouder, Joëlle Pellet

Avant propos



Dès la troisième édition, les conférences TOTh ont trouvé une structuration qui traduit bien à la fois le caractère scientifique et pluridisciplinaire de la terminologie et l'intérêt de notre communauté pour d'autres domaines partageant des préoccupations communes.

Ainsi, la conférence d'ouverture a été donnée par une personnalité invitée issue d'une discipline différente de la nôtre – ici la phylogénèse – mais pour laquelle le langage et la pensée jouent également un rôle primordial.

Les contributions se sont réparties naturellement, et par le jeu des évaluations de façon équitable, en trois groupes ayant donné lieu à trois sessions.

Le premier groupe a rassemblé les articles portant principalement sur la dimension linguistique de la terminologie. Ont été abordés l'extraction terminologique à partir de dictionnaire, la place accordée aux corpus dans la construction de terminologies, l'acquisition de connaissances à partir de textes et l'apport des ressources linguistiques issues du web.

La deuxième session s'est donc logiquement intéressée à la dimension conceptuelle de la terminologie. Les notions de concept, de relation, d'ontologie ont été au cœur des présentations portant sur les cartes conceptuelles pour les bibliothèques numériques, les relations dynamiques et les graphes conceptuels, l'alignement d'ontologies et l'accès multilingue aux ontologies.

Enfin, la troisième session a été consacrée à la présentation de plusieurs applications terminologiques pour des secteurs aussi différents que l'ingénierie nucléaire, l'informatique, le domaine bancaire ou l'agriculture biologique. Il est à souligner que ces applications ont permis d'aborder différents points théoriques tels que la variation terminologique, la diachronie ou la structure des dictionnaires.

La richesse des débats qui ont animé ces deux jours de conférence – chaque présentation, questions comprises, s'est vue allouer plus de quarante cinq minutes de temps de parole – a été certainement une des plus belles récompenses pour les participants de TOTh 2009.

Christophe Roche

Président du Comité Scientifique

Table des matières

CONFERENCE INVITEE

<i>La nomenclature biologique aujourd'hui : que reste-t-il de Linné ?</i>	1
Michel Laurin	

SESSION 1

<i>Approche lexico-sémantique de l'extraction terminologique : utilisation de ressources lexicographiques et validation sur corpus</i>	19
Bertrand Gaiffe, Evelyne Jacquey, Laurence Kister	
<i>Quelle place accorder aux corpus dans la construction d'une terminologie ?</i>	33
Marie Calberg-Challot, Pierre Lerat, Christophe Roche	
<i>Extraction de connaissances orientées évolution dans les textes techniques</i>	53
Kata Gabor, François Rousselot, François De Bertrand de Beuvron	
<i>Corpus et Web : deux alliés pour la construction de l'enrichissement automatique de classes conceptuelles</i>	73
Nicolas Béchet, Mathieu Roche, Jacques Chauché	

SESSION 2

<i>Following the path between conceptual maps and visual thesauri</i>	93
Olga Bessa Mendes	
<i>Dynamic concept relations: a definition and representation proposal</i>	107
Chiara Messina	
<i>Construction et alignement d'ontologies pour évaluer le risque alimentaire</i>	127
Liliana Ibanescu, Patrice Buche, Juliette Dibie-Barthélemy	
<i>Accès multilingue à une ontologie par des correspondances avec un lexique pivot</i>	143
David Rouquet, Hong-Thai Nguyen	
<i>La reformulation : processus dynamique d'acquisition des connaissances. Le cas du discours technique arabe d'Internet</i>	161
Andrée Affeich	

SESSION 3

<i>Structuration d'un dictionnaire de spécialité pour sa publication sur internet. Bénéfices du langage XML</i>	181
Jacques Joseph	
<i>Mémoire du Club informatique des grandes entreprises françaises (CIGREF) : nouveau plan de classement</i>	197
Jean-Yves Gresser, M.P. Lacroix	
<i>Les secteurs d'activité à l'épreuve du discours</i>	217
Frédéric Erlos	
<i>De l'agriculture biologique aux espaces naturels : une étude des syntagmes terminologiques à l'intérieur des textes de spécialité</i>	235
Elisa Lavagnino	
<i>Pages blanches</i>	253

Le projet NucSTML

Structuration d'un dictionnaire de spécialité en vue de sa publication sur internet

Bénéfices du langage XML

Jacques JOSEPH

Résumé : Le langage XML est particulièrement bien adapté à la représentation d'un dictionnaire de spécialité destiné à être diffusé sur le Web. La simplicité du langage assure l'interopérabilité et la portabilité des documents et rend possible leur diffusion sur une grande variété de systèmes d'exploitation. Le format est facile à mettre en œuvre, à installer sur un serveur, il est économe en ressources. Il sépare le fonds de la forme, facilitant ainsi la gestion du contenu et assurant la pérennité du document. Les feuilles de styles et les langages associés permettent de formuler des requêtes puissantes et des fouilles profondes et rapides dans l'arborescence du fichier source, il peut être une aide à l'élaboration de réseaux conceptuels. L'application présentée dans cet article fait suite aux travaux de Marie Calberg pour l'élaboration d'un dictionnaire de l'ingénierie nucléaire (Calberg-Challot *et al.* 2008a) et décrit sa mise en œuvre au sein d'une grande entreprise du secteur de l'énergie. Le projet NucSTML (Nuclear Science and Technology Markup Language) a pour objet de réaliser un modèle de document simple, adapté à la publication et à l'échange de données linguistiques dans le domaine des sciences et techniques de l'énergie nucléaire.

Mots-clés : Modèle, Dictionnaire, Nucléaire, Internet, XML.

1. Introduction

Pour faire face à l'évolution des techniques et du vocabulaire associé, à l'élargissement de leur périmètre et à l'arrivée d'une nouvelle génération de femmes et d'hommes dans les domaines techniques, commercial ou administratif, ainsi qu'aux besoins de la société, certains acteurs de l'industrie et de la recherche nucléaire française ont entrepris de mettre à jour leurs principaux dictionnaires techniques voire d'en créer de toute pièce. Le Commissariat à l'Énergie Atomique (Cea) a publié en 2008 sous la conduite de M. le Haut Commissaire et avec la collaboration active de ses principaux partenaires - Autorités de Sûreté Nucléaire (ASN), Institut de Recherche sur la Sûreté Nucléaire (IRSN), Areva, Edf, Andra - notamment, le *Dictionnaire des sciences et techniques nucléaires* (Bigot *et al.* 2008), regroupant près de quatre mille entrées qui couvrent l'ensemble des secteurs : scientifique et industriel, civil ou de défense, ce dictionnaire constituant la mise à jour d'un ouvrage plus très ancien (Dictionnaire des sciences et techniques nucléaires). Areva-Np a de son côté entrepris de réaliser son propre dictionnaire de spécialité (Calberg-Challot *et al.* 2008a), (Calberg-Challot *et al.* 2008b) dédié à l'ingénierie nucléaire. Notons que le Cea, répondant à sa mission d'information, a diffusé l'ouvrage de façon publique, tandis que dans le cas des industriels, le contenu reste prioritairement réservé à un usage interne. La caution active des plus hautes autorités prouve l'importance qu'elles attachent à la maîtrise de la terminologie technique et à sa diffusion.

Cet article décrit une expérience industrielle transposable, présente une méthode pour réaliser de façon économique et efficace un dictionnaire de spécialité, facilement intégrable dans un site internet ou intranet, dans des applications de gestion documentaire, portable et gérable de façon simple et pérenne.

2. Analyse de la diffusion sur internet de la terminologie nucléaire

De nombreux organismes français ou étrangers, (lexique nucléaire de A à Z), des sociétés savantes telles que la Société Française de l'Énergie Nucléaire (SFEN), des électriciens francophones (HydroQuébec), des organismes de régulation (Glosary) ou des organismes internationaux tels que l'Agence Internationale pour l'Énergie Atomique (AIEA 2006), des laboratoires de protection (Lexique Cea), (Santé Canada : glossaire), proposent dans leur site internet, sous la dénomination *lexique* ou *glossaire*, des listes de termes techniques et scientifiques en rapport avec leur activité

et leurs métiers (production d'énergie, physique des réacteurs, protection des populations, biologie, médecine, contrôle, instrumentation...).

Ces listes sont généralement construites sur un modèle similaire : un contenu de termes communs tels que *actinides*, *ALARA*, *Becquerel*, *uranium* (physique fondamentale) ou de termes spécifiques *pressuriseur*, *réacteur* (domaine des réacteurs), *évacuation* (domaine de la protection), etc... On note deux méthodes de diffusion électronique : documents PDF (Postscript Document Format) de manipulation peu aisée, pages HTML (Hyper Text Markup Language) d'utilisation plus souple et mieux adaptée à une diffusion électronique, mais avec les limitations propres à ce langage (nombre de balises limité, pauvre d'un point de vue sémantique, mélange de la forme et du fonds...)

3. Historique et justification du choix du langage XML

3.1. Rappel des principales caractéristiques du langage

XML est un métalangage de balises, permettant de représenter et de stocker des documents structurés dans un format texte, de les partager, de les transformer en tout format compatible les rendant facilement publiables sur le Web. Les applications justifiant l'utilisation de XML sont variées et multiples : oeuvres poétiques ou théâtrales (Gooss 1999), (Jung 2000), mise en place de systèmes documentaires à caractère administratif par le gouvernement du Québec (Marcoux 2001), ontologie (Gandon 2002). Les DTD (**D**ocument **T**ype **D**efinition) permettent de créer des modèles de document auquel peuvent se référer différents groupes travaillant sur une thématique similaire. Par exemple l'initiative libre Text Encoding Initiative (TEI) préconise un standard de représentation des types de textes relevant d'une représentation XML (TEI).

3.2. Les dictionnaires et XML justification des choix

La publication en ligne de dictionnaires généraux ou techniques invoquant la technologie XML se développe dans de nombreux secteurs industriels ou de recherche. Dans la banque ou le bâtiment par exemple où l'interopérabilité, la standardisation des échanges, l'indépendance vis-à-vis des systèmes d'exploitation sont les arguments principaux (Henriet *et al.* 2006), (Méta modèle de Dictionnaire Standard de la Construction) en faveur de son usage. Pour d'autres, XML est : un moyen puissant de hiérarchisation et de structuration des données textuelles (Bernard *et al.*),

une méthode de balisage de dictionnaires anciens (Rey 2004) ou bien encore un moyen utilisé dans l'élaboration de dictionnaires multilingues (Mangeot *et al.* 2003), d'autres encore soulignent que l'accès à n'importe quelle partie d'un dictionnaire, grâce à XML amplifie la capacité de diffusion des versions électroniques d'un dictionnaire (Elchacar 2008). Notons enfin la standardisation de dictionnaires multilingues grâce à XML (Maks *et al.* 2008).

4. Construction du document XML

La méthode de collecte et de sélection des entrées, ainsi que la formulation des définitions adoptées par Areva-Np a été décrite par Marie Calberg (Calberg-Challot *et al.*), le fonds terminologique repose sur des ouvrages techniques, cf. par exemple (Coppolani *et al.* 2004). Un nombre limité de descripteurs, a été sélectionné, certains sont obligatoires d'autres facultatifs. Les données fondamentales sont considérées comme **élément : dictionnaire** (élément racine), *entry* (élément courant) contenant les éléments fils : *expression*, *équivalent(s) anglais*, *définition*, d'autres facultatifs : *note(s)*, *exemples*, *renvoi(s)*, *forme abrégée*, *synonyme(s)*, *illustration*. Les données complémentaires sont considérées comme **attribut** (ex : *langue*, *date* -de validation de l'entrée-, *domaine*, *auteur*, *statut* -approuvé-) : Chaque élément *entry* possède un identificateur unique et obligatoire sous forme d'attribut, *id*, un nom de *domaine*, chaque élément référence (*ref*) possède un attribut *idref* qui assure la relation avec la cible repérée par son identificateur. L'arborescence correspondante du dictionnaire est représentée sur la Figure 1.

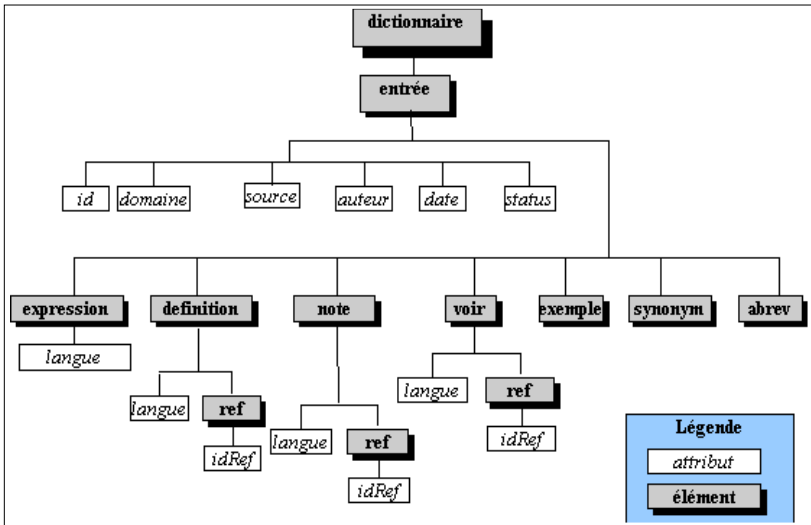


Figure 1. Arborescence général du dictionnaire

Le Tableau 1 représente le contenu de l'élément de base du document, la DTD est représentée dans le Tableau 2.

Le fichier XML est créé à partir du tableur source, à l'aide de langages de programmation classiques (Visual basic ou Php), qui permettent de transformer le contenu et de l'intégrer entre les balises. Simultanément la liste de l'ensemble des termes présents dans le dictionnaire est insérée dans un module *javascript*, servant de support à un formulaire de requête.

```

<entry id='dinId5' langue='fr' domain='P' resp='RCC-P' date='2009-02-15' status='GPA'>
<expression langue='fr'>accident</expression>
<expression langue='en'>accident</expression>
<synonym>transitoire accidentel</synonym>
<definition langue='fr'>Évènement hautement improbable, pouvant entraîner
l&apos;endommagement d&apos;une ou plusieurs <ref idref='dinId112'>barrières de confinement</ref>
avec éventuel relâchement de produits radioactifs et demandant
la mise en service des <ref idref='dinId1015'>systèmes de sauvegarde</ref> par
le <ref idref='dinId1005'>système de protection</ref>.</definition>
<note langue='fr'>Deux catégories d&apos;accidents sont prises en compte :
les accidents de fréquence très faible dits de condition 3,
et les accidents hypothétiques dits de condition 4.</note>
<voir><ref idref='dinId949'>situation de troisième catégorie</ref>,
<ref idref='dinId948'>situation de quatrième catégorie</ref></voir>
</entry>

```

Tableau 1. Exemple d'élément du dictionnaire

```

<!ELEMENT entry (illustration?|expressions+)>
<!ELEMENT expression (sub|sup)*>
<!ELEMENT definition (ref|sub|sup)*>
<!ELEMENT note (ref|sub|sup)*>
<!ELEMENT abbrev (ref|sub|sup)*>
<!ELEMENT voir (ref|sub|sup)*>
<!ELEMENT exemple (ref|sub|sup)*>
<!ELEMENT ref (sub|sup)*>
<!ATTLIST entry id ID #REQUIRED>
<!ATTLIST entry langue CDATA #IMPLIED>
<!ATTLIST entry categorie CDATA #IMPLIED>
<!ATTLIST entry resp CDATA #IMPLIED>
<!ATTLIST entry status CDATA #IMPLIED>
<!ATTLIST entry date CDATA #REQUIRED>
<!ATTLIST illustration source CDATA #REQUIRED>
<!ATTLIST illustration alt CDATA #IMPLIED>
<!ATTLIST expression language (fr|en) #REQUIRED>
<!ATTLIST definition language (fr|en) #REQUIRED>
<!ATTLIST note language (fr|en) #REQUIRED>
<!ATTLIST ref idref IDREF #REQUIRED>
<!ENTITY din SYSTEM "din.xml">
    
```

Tableau 2. DTD du document XML

4.1. Balisage des références.

XML permet d'organiser les renvois en marquant les termes cible, les instructions XSL se chargeant d'en assurer l'affichage de façon automatique, dans notre cas un renvoi est marqué par l'élément `<ref>..</ref>`, l'attribut *idref* a la valeur de l'identificateur du terme cible :

```
<ref idref='idValue'>mot cible</ref>
```

Cette méthode permet d'afficher de façon automatique les termes associés au moment de la requête sur le poste client, à l'aide de "templates" placés dans les feuilles de style XSL dont un exemple est représentée dans le Tableau 3. Le marquage peut être réalisé de façon automatique à l'aide d'algorithmes, soit au début du processus à l'aide de langages de programmation (Vb, Php), soit en fin de processus en utilisant les instructions XSL.

```

<xsl:template match='ref'>
<xsl:if test="//entry[@id=$idreference]">
<a><xsl:attribute name="href">
<script>definition('&apos;
<xsl:value-of select="@idref"/>&apos;)</xsl:attribute>
<xsl:apply-templates/>
</a>
    
```

```
<xsl:if>
<xsl:template>
```

Tableau 3. Type de modèle XSL utilisé pour relier une référence à sa cible

4.2. Requêtes

L'association des langages, *HTML*, *javascript*, *XSL*, *XPATH*, permet de recueillir les arguments de la requête formulée à partir du poste client par l'intermédiaire d'un formulaire et de la transmettre par l'entremise d'un script *java script* sous forme de paramètres (Tableau 4) à la feuille de style *XSL* (Tableau 5), les instructions de la feuille de style organisent la requête à l'aide d'instructions *XPATH*, qui pointent les éléments recherchés et les retournent au poste client au format *HTML*. La page d'interface *HTML*, propose à l'utilisateur plusieurs manières de parcourir le dictionnaire pour afficher les définitions : soit directement à partir de la liste dynamique de l'ensemble des entrées, soit à partir d'une chaîne de caractères figurant dans le contenu de l'entrée (*expression*, *définition*, *notes*, *exemples*).

```
var xslDoc = new ActiveXObject("Msxml2.FreeThreadingDOMDocument");
var xmlDoc = new ActiveXObject("Msxml2.DOMDocument");
xslDoc.async = "false";
xslDoc.load("dictionnaire.xsl");
xmlDoc.validateOnParse = false;
xmlDoc.async = false;
xmlDoc.load("dictionnaire.xml");
var xslt = new ActiveXObject("Msxml2.XSLTemplate");
xslt.stylesheet = xslDoc;
var xslProc = xslt.createProcessor();
xslProc.input = xmlDoc;

function definition(arg,lang,option) {
    xslProc.addParameter("langue",lang);
    xslProc.transform();
    xslProc.addParameter("terme",arg);
    xslProc.transform();
    xslProc.addParameter("ident","");
    xslProc.transform();
    zoneDefinition.innerHTML=xslProc.output;
}
```

Tableau 4. Script d'interface transmettant la valeur des paramètres de la requête à la feuille de style

Les instructions de traitement des chaînes de caractères de *XPATH* permettent d'identifier l'élément concerné, un exemple est indiqué ci-dessous. L'expression présentée, applique un modèle (templates) à toutes les expressions possédant l'attribut langue (@langue) = \$langue) et qui sont égales au terme (\$terme).

```
<xsl:apply-templates  
select="//expressions[expression[@langue=$langue]=$terme]"/>
```

Le contenu de l'entrée est affiché et les termes associés (renvois) sont repérés et marqués, grâce à la correspondance *idref* = *id*, leur définition complète est affichée sur demande de l'utilisateur, l'image illustrant le terme s'affiche lorsqu'elle existe, la demande étant traitée comme une nouvelle requête suivant la même méthode. La Figure 2 présente le processus général de création et d'exploitation du dictionnaire électronique.

```
xmlns:xdt="http://www.w3.org/2005/xpath-datatypes">  
<xsl:output method="html" version="1.0"  
  encoding="ISO-8859-1" indent="yes"/>  
<!--Liste des paramètres transmis par javascript-->  
<xsl:param name="terme"/>  
<xsl:param name="langue"/>  
<xsl:param name="iopt"/>  
<!--=====-->  
<xsl:template match="/">  
<xsl:apply-templates/>  
<!--Intégration des compléments de styles-->  
</xsl:template>  
<xsl:include href="dictionnaire3.xsl"/>  
</xsl:stylesheet>
```

Tableau 5. Feuille de style. Recueil de la valeur des paramètres transmise par la fonction *definition* ()

4.3. Diffusion du dictionnaire

La méthode est utilisée pour représenter plusieurs dictionnaires de spécialité. La Figure 2 présente le processus général de création et d'exploitation du dictionnaire électronique.

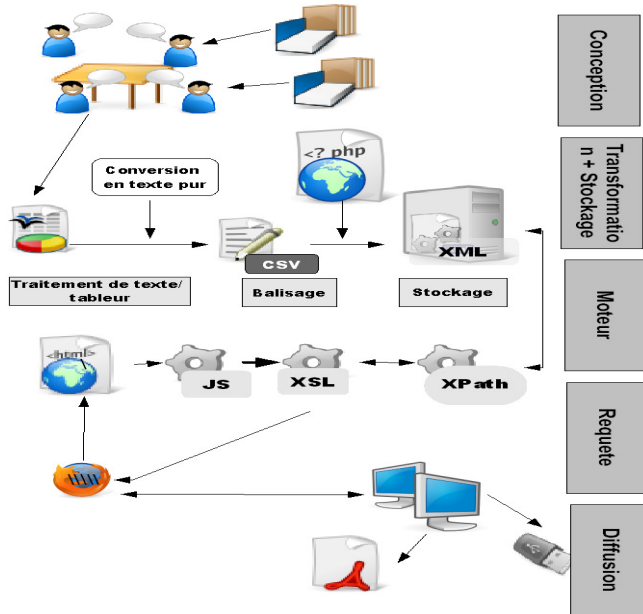


Figure 2. Processus général de production et de diffusion du dictionnaire

4.4. Facilité de partage et d'association, le dictionnaire virtuel

Une des bénéfices les plus fructueux de XML réside dans la possibilité de constituer un document unique à partir de plusieurs documents épars obéissant à la même DTD. Ceci permet par exemple de créer un dictionnaire virtuel multilingue par assemblage de dictionnaires ou de lexiques monolingues. Pour cela il faut créer un document XML, qui devient l'élément père et dont chaque fils est l'un des dictionnaires, ceci est matérialisé par une nouvelle DTD (Tableau 6) invoquée dans le document XML virtuel (Tableau 7). Le schéma d'ensemble de la méthode est représenté sur la Figure 3.

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<?xml-stylesheet type="text/xsl" href="dictionnaire.xsl"?>
<!DOCTYPE dictionnaireVirtuel SYSTEM "dictionnaireVirtuelDt
<dictionnaireVirtuel>
&dictionnaire_1;
&dictionnaire_2;
&dictionnaire_3;
</dictionnaireVirtuel>
    
```

Tableau 6. DTD du dictionnaire virtuel.

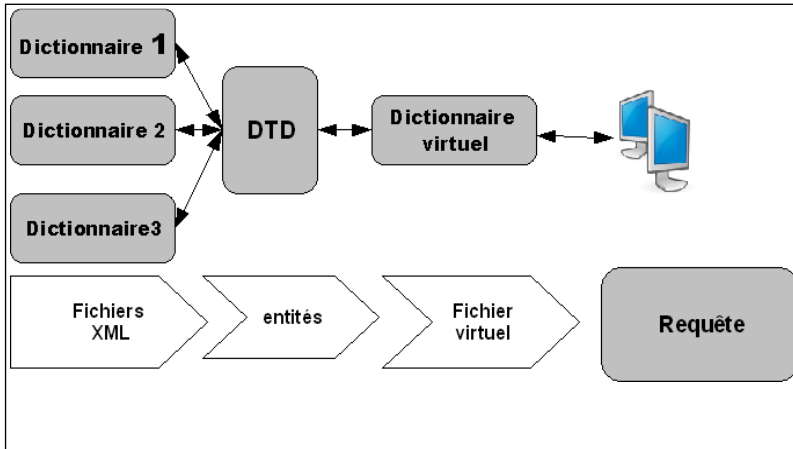


Figure 3. Processus de création d'un dictionnaire virtuel à partir de dictionnaires élémentaires

5. Extension des applications

5.1. Recherche d'associations de termes

La structuration du dictionnaire suivant un schéma XML, en dehors de la publication du contenu sous la forme que nous venons de présenter, permet d'en exploiter le contenu pour rechercher des associations et des groupements de termes, soit définis a priori (renvois), soit implicitement. Pour cela nous avons rédigé une feuille de style XSL, qui par la simple utilisation de *templates* permet d'extraire des informations souhaitées. Nous avons réalisé un test qui a porté sur l'entrée *système de sauvegarde* et sur le terme isolé *sauvegarde* considéré comme une chaîne de caractères, la notion de *système de sauvegarde* ayant été analysée par Marie Calberg-Challot (Calberg-Challot *et al.* 2008a) pour en dégager l'aspect conceptuel.

5.2. Termes explicitement associés

Nous avons explicité plusieurs couches de termes associés en cascade (renvois dans les renvois...), en exploitant les éléments `<ref>` par le biais de l'attribut `idref`. Ceci permet par exemple d'observer le maillage formé autour de l'expression *système de sauvegarde* et de vérifier si les termes associés peuvent relever du concept. Nous nous sommes limités à trois couches successives, le résultat est indiqué dans le Tableau 8. Ceci révèle par exemple les termes *source froide*, *chaleur résiduelle*, *sous critique* qui ne sont pas des objets relevant des *systèmes de sauvegarde* mais qui cependant en conditionnent fortement la conception et le dimensionnement, le terme

chaleur résiduelle par exemple est mentionné dans le graphe du réseau conceptuel de «système».

1 - accident hors dimensionnement
2 conditions de fonctionnement complémentaires
3 arrêt d'urgence
3 générateur de vapeur
3 source froide
1 - fonction de sauvegarde
2 arrêt d'urgence
3 système d'arrêt d'urgence
3 système de protection
2 arrêt sûr
3 chaleur résiduelle
3 sous-critique
3 tranche nucléaire
2 bâtiment réacteur,
2 chaudière nucléaire

Tableau 8. Liste des 3 couches de renvois successifs dans l'expression système de sauvegarde

5.3. Associations implicites

En effectuant une recherche à partir de la chaîne de caractères "*sauvegarde*" nous étendons le maillage du document, le résultat est illustré dans le Tableau 9. Les entrées mentionnées contiennent au moins une occurrence de la chaîne de caractères dans l'un quelconque des éléments fils. Nous retrouvons dans ce tableau des termes déjà mentionnés dans la première liste, ainsi que certains termes relevant du concept *système de sauvegarde* mentionnés dans la publication en référence tels que : *aspersion d'enceinte, système d'injection, système de refroidissement, alimentation de secours*. Une analyse plus détaillée est cependant nécessaire pour valider les relations qui existent avec le concept *système de sauvegarde* et les expliciter, le terme *période longue* par exemple fait bien référence à la *fonction de sauvegarde*. La méthode permet de clarifier a posteriori la cohérence du contenu du dictionnaire, le terme *sauvegarde* est lié aux termes *accident, protection, refroidissement, secours, enceinte de confinement, sécurité, auxiliaires, injection*. Les *systèmes* qui apparaissent dans la liste concordent avec ceux qui apparaissent dans le graphe de Marie Calberg-Challot. L'analyse du terme *système de traitement des effluents primaires* montre une concordance partielle avec le réseau conceptuel, dans lequel le

terme *bore* par exemple n'est pas mentionné, bien que cet élément utilisé dans le circuit primaire comme absorbant de neutrons, contribue fortement à produire des effluents liquides.

Les deux méthodes s'apparentent plus à une extraction de termes dans un document qu'à la recherche de sens et de liens, elles montrent cependant le parti que l'on peut tirer d'une structuration avec XML qui permet la fouille très fine du contenu et par exemple de dégager un réseau de mots dont certains peuvent après analyse, se référer à un concept identique.

<p>cident</p> <ul style="list-style-type: none"> - accident de dimensionnement - accident enveloppe de perte de réfrigérant primaire - actionneur de sauvegarde - automate de protection - auxiliaires de sauvegarde - auxiliaires électriques - auxiliaires secourus - bâtiment des auxiliaires de sauvegarde - bâtiment des auxiliaires nucléaires - circuit d'eau brute secourue - classe de sûreté 1E - fonction de sauvegarde - période longue 	<ul style="list-style-type: none"> - système d'alimentation de secours des générateurs de vapeur - système d'aspersion de l'enceinte de confinement - système d'eau brute secourue - système d'injection de sécurité - système de contrôle volumétrique et chimique - système de protection - système de protection intégré numérique - système de refroidissement intermédiaire - système de sauvegarde - système supports - train de sauvegarde - unités logiques de sauvegarde - voies de sauvegarde redondantes
---	---

Tableau 9. Liste des entrées extraites dans l'ensemble du dictionnaire contenant au moins une fois la chaîne de caractère "sauvegarde"

6. Bénéfices d'XML

Nous pouvons souligner quelques-uns des avantages majeurs qui à nos yeux procurent une grande facilité d'installation dans un système industriel. La construction d'un dictionnaire fondée sur XML peut être réalisée au fur et à mesure de l'avancement des travaux d'élaboration du dictionnaire, directement dans les groupes de travail.

Les avantages procurés par cette méthode sont donc nombreux :

- développements informatiques légers et autonomes ;
- portabilité ;
 - o le dictionnaire peut être transporté par des moyens légers (ordinateurs portables, clés usb, Cd Roms),
 - o possibilité d'extension et d'intégration dans des systèmes d'information plus étendus,

- flexibilité et évolutivité,
 - o les modifications ultérieures ne nécessitent pas de revoir la forme,
 - o facilité de maintenance,
- pérennité des supports ;
 - o le format texte est universel,
 - o la forme peut être adaptée et évoluer suivant les besoins sans avoir à modifier le contenu,
- mise en commun de dictionnaires complémentaires par exemple dans un univers professionnel éclaté ou multilingue ;
- exploitation ultérieure en utilisant la structure du document ;
 - o recherche de similitudes,
 - o aide à la recherche de concepts.

7. Conclusions

Le langage XML et ses langages associés permettent de représenter et de partager simplement et sans outils sophistiqués, le contenu d'un dictionnaire. Le modèle de document NucSTML reproduit la structure d'un dictionnaire de spécialité utilisable par tous les protagonistes de l'industrie nucléaire. Un bloc note et un navigateur permettent d'en stocker et d'en diffuser le contenu sur tout système portable notamment, de plus XML autorise la mutualisation de ressources linguistiques éparses, sans obligation de centralisation, permettant ainsi à différents groupes linguistiques appartenant à la même communauté technique de construire séparément un dictionnaire de spécialité monolingue et de créer un dictionnaire virtuel regroupant chacun des dictionnaires au moment de la requête sur le poste client. Un dictionnaire conçu suivant les principes XML rend possible son intégration dans des systèmes de GED ou de KM. Le choix initial de la structure du document et la construction de groupes reconnus par des attributs spécifiques devraient permettre d'aller au delà d'une simple représentation séquentielle de mots et faciliter des regroupements ou des associations de mots.

Bibliographie

Barraclough I. (2006) : *IAEA safety glossary. Terminology used in nuclear, Radiation, radioactive waste and transport safety*, Version 2.0., Departement of nuclear safety and security, IAEA, Vienna, Austria

[Lexique nucléaire de A à Z. <http://www.areva-np.com/scripts/info/publigen/content/templates/show.asp?P=882&L=FR>

Bernard Pascale, Dendien Jacques, Pierrel Jean-Marie *A computerized dictionary : Le Trésor de la langue française informatisé (TLFI)*. ATILF. UMR 7118-CNRS/Université Nancy

Calberg M. Dumont X. (2004) : *Projet du dictionnaire de l'ingénierie nucléaire*, Note interne, Areva-NP

Calberg-Challot Marie, Candel Danielle, Roche Christophe (2008a) : *De la variation des usages au consensus terminologique : vers un dictionnaire de l'ingénierie nucléaire*, Conférences Toth. Annecy, 5-6 juin 2008

Calberg-Challot Marie, Danielle Candel, Didier Bourigault, Xavier Dumont, et Jacques Joseph (2008b) : *Une analyse méthodique pour l'extraction terminologique dans le domaine du nucléaire*, Terminology 14 :2, pp. 183-203

Dictionnaire des sciences et techniques nucléaires (1975) : 3^{ème} édition. Eyrolles. Paris

<http://www.cadarache cea.fr/fr/accueil/lexique.php>

Santé canada : Glossaire. <http://www.bc-sc.gc.ca/ed-ud/event-incident/radiolog/info/glossary-glossaire-fra.php>

Coppolani Pierre, Hassenboehler Nathalie, Joseph Jacques, Pétetrot Jean-François, Py Jean-Pierre, Zampa Jean-Sébastien (2004) : *La chaudière des réacteurs à eau sous pression*, Collection Génie Atomique, Instn, EDP Sciences

Méta modèle de Dictionnaire Standard de la Construction. <http://www.edibatec.org/Accueil/SDC/ProjetSDC.htm>

Elchacar Mireille *Le projet "FRANQUS" : la langue française vue par le Québec*, Projet de Thèse de doctorat, Université de Sherbrooke, Québec, Ca

Fabien Gandon (2002) : *Ontology Engineering a Survey and a Return on Experience*. *Rapport de Recherche*, Ch 4. Pg 61. ACACIA team. N° 4396

Gossens Michel (1999) : *XML et XSL : un nouveau départ pour le web*. Cahiers GUTenberg n° 33-34 _ Congrès GUT99-Journée XML

Henriet Laurent, Fagnent Sylvain (2006) : *Standardisation des échanges : Mise en œuvre d'un dictionnaire bancaire d'interopérabilité*, Revue Banque, N° 679

Hydroquébec, Glossaire des termes nucléaires, <http://www.hydroquebec.com/fr/>

XML, Backgrounder. *Technology and applications*.

Software AG (2000) : The XML Company, Darmstatt, De

Maks Isa, Tiberius Carole, van Veenendaal Remco. *Standardising bilingual lexical resources according to the Lexicon Markup Framework*, Dutch HLT agency (TST-centrale), Institute for Dutch Lexicology (INL), Leiden, Netherlands

Mangeot-Nagata Mathieu, Bilac Slaven (2004) : *Construction collaborative d'un dictionnaire multilingue, Le projet Papillon*. Actes de JSF 2003, National Olympic Memorial Youth Center, Tokyo, Japon

Manuélian Hélène et Schang Emmanuel (2007) : *XML, DTD et TEI pour un dictionnaire étymologique des créoles*, Universités de Cergy Pontoise et Orléans

Marcoux *et al.* (2001) : *XML en route au gouvernement du Québec*, Groupe de Recherche sur les documents structurés, École de Bibliothéconomie et des Sciences de l'Information, Université de Montréal, Rapport Final

Glossary <http://www.nrc.gov/reading-rm/basic-ref/glossary.HTML>

Rey Christophe, Zaouy Corinne (2004) : *Balisage XML "ciblé" : une nouvelle approche dans l'informatisation des corpus*, Université de Provence, Colloque international sur la fouille des textes, La Rochelle

Bigot B., Santarini G. *et al.* (2008) : *Dictionnaire des sciences et techniques nucléaires, 4e édition*, OmniSciences

Text Encoding Initiative, The XML version of the TEI Guideline, Chapitre 13, The terminological database

A propos des auteurs

Jacques Joseph

45 rue de Saint Cyr

69009 Lyon

jacquespjoseph@aol.com