

Rubrique préparée par Fiammetta Namer

*Université Nancy2 de Nancy, UMR « ATILF »
Fiammetta.Namer@univ-nancy2.fr*

Javier COUTO (jcouto@fing.edu.uy)

Titre : Une plate-forme informatique de Navigation Textuelle : modélisation, architecture, réalisation et applications de NaviTexte

Mots-clés : navigation textuelle, représentation informatique des textes, modélisation de connaissances, plate-forme informatique de navigation textuelle.

Title : *A text navigation system : modelling, architecture, development and applications of NaviTexte.*

Keywords : *text navigation, text representation, knowledge modelling text navigation system.*

Thèse de doctorat en Mathématiques, Informatique et Applications aux Sciences de l'Homme, Université de Paris-Sorbonne (Paris IV), LaLICC, Langages, Logiques, Informatique, Cognition et Communication, UMR 8139 du CNRS, sous la direction de M. Jean-Luc MINEL, Ingénieur de Recherche (HDR), CNRS. Soutenue le 10/07/2006.

Jury : M. Patrice Enjalbert (Pr, Université de Caen et GREYC, rapporteur et président), M. Jean-Luc Minel (IR HDR, LaLLIC-CNRS, directeur), M. Jean-Paul Sansonnet (DR., LIMSI-CNRS, rapporteur), M. Yves Jeanneret (Pr, Université de Paris-Sorbonne & CELSA, examinateur), Mme Lita Lundquist (Pr, CBS, Danemark, examinatrice), Mme Dina Wonsever (Pr, Universidad de la República, Uruguay, examinatrice).

Résumé : *Au long de l'histoire, des instruments de recherche d'information ou d'aide à la lecture, fondés sur la notion de page, tels que la table des matières, les index, les renvois, etc., ont été introduits. Dès l'arrivée de l'informatique, ces instruments, de nature typiquement statique, se sont multipliés et ils ont vu croître leur puissance. Dans le cadre général de l'histoire du texte numérique, l'hypertexte place un jalon du point de vue conceptuel, et l'utilisation massive d'Internet a répandu son utilisation à grande échelle.*

Le terme de navigation textuelle reçoit de multiples interprétations, la plus commune renvoyant inévitablement au processus mis en œuvre par les outils de

navigation utilisés pour circuler dans les documents hypertextes. Néanmoins, notre conception de la navigation textuelle se démarque de la navigation hypertextuelle traditionnelle car nous considérons que circuler ou naviguer dans un texte est l'expression d'un processus cognitif qui convoque des connaissances qui sont propres à la finalité de la navigation. Nous formulons l'hypothèse que la démarche du lecteur peut être assistée par l'exploitation de connaissances, présentes dans les textes, qui peuvent être, en partie, modélisées sous une forme déclarative. Autrement dit, il ne suffit pas de créer des liens mais il est nécessaire d'explicitier l'opération de navigation. De plus, ce processus de définition d'opérations de navigation doit être mis en œuvre par un « expert » capable d'encoder ces connaissances.

Ce travail de thèse présente quatre contributions principales. En premier lieu, une représentation des textes spécifique à la navigation textuelle est définie. En deuxième lieu, un langage formel de modélisation des connaissances de visualisation et de navigation, nommé Sextant est proposé. Les constructions possibles du langage sont données par une syntaxe. Le sens des constructions syntaxiques du langage Sextant est déterminé par une sémantique opérationnelle. En troisième lieu, une plate-forme logicielle dédiée à la navigation textuelle, nommée NaviTexte, a été implémentée. Dans cette implémentation, trois choses ont été développées : une représentation informatique des textes spécifique à la navigation textuelle, fondée sur la proposition d'un encodage XML des textes ; un interpréteur d'une version réduite de Sextant, le langage de modélisation des connaissances, fondé sur la proposition d'un encodage XML de ce langage ; un environnement capable de traiter les textes, d'interpréter le langage Sextant et de gérer l'interaction avec l'utilisateur. En dernier lieu, diverses applications de la plate-forme logicielle NaviTexte à des cas réels d'utilisation ont été mises en œuvre.

URL où la thèse pourra être téléchargée :

1. http://www.lalic.paris4.sorbonne.fr/PUBLICATIONS/affiche_theses.php?id=41
2. [http://hal.archives-ouvertes.fr/docs/00/08/76/06/PDF/These_Couto_-_Paris_Sorbonne_\(Paris_IV\)_-2006.pdf](http://hal.archives-ouvertes.fr/docs/00/08/76/06/PDF/These_Couto_-_Paris_Sorbonne_(Paris_IV)_-2006.pdf)

M. Jorge GARCÍA-FLORES (jorge.gflores@paris4.sorbonne.fr)

Titre : Annotation sémantique des spécifications informatiques de besoins par la méthode d'Exploration Contextuelle : une contribution des méthodes linguistiques aux conceptions de logiciels

Mots-clés : spécification de besoins, annotation, sémantique, conception de logiciels, texte, action, contrôle, finalité, contrainte, sémantique, analyse discursive, exploration contextuelle, verbes d'action, marqueur linguistique, système

d'annotation automatique EXCOM.

Title : *Semantic annotation of software requirements specifications with the Contextual Exploration Method : a contribution of linguistic methods to the requirements engineering process.*

Keywords : *linguistics, annotation, requirements engineering, automatic extraction, action, control, goal, constraint, Contextual Exploration Method, discourse analysis, action verbs, linguistic marker, semantic annotation, EXCOM semantic annotation system.*

Thèse de doctorat en Informatique Linguistique, Université Paris-Sorbonne (Paris IV), LaLICC, Langages, Logiques, Informatique, Cognition et Communication, UMR 8139 du CNRS, sous la direction de M. Jean-Pierre Desclés, Professeur. Soutenue le 15/01/2007.

Jury : Mme Colette Rolland (Pr, Université Panthéon-Sorbonne & Centre de Recherche en Informatique, présidente), M. Jean-Pierre Desclés (Pr, Université Paris-Sorbonne & LaLICC, directeur), Mme María Teresa Cabré Castelleví (Pr, Université Pompeu Fabra, Barcelone, rapporteur), M. Jean-Paul Sansonnet, (Pr, LIMSI-CNRS, rapporteur), M. Brahim Djoua (MC, Université Paris-Sorbonne, examinateur), M. Yves Marcoux, (Pr, Université de Montréal, examinateur).

Résumé : *La spécification de besoins est une activité fondamentale dans l'ensemble des méthodes de conception et mise en œuvre des systèmes informatiques. Cette thèse présente une méthode d'annotation sémantique des textes de Spécification de Besoins Informatiques (SBI) pour assister la conception de logiciels ; une méthode dont les annotations, de nature sémantique, peuvent être exploitées pour l'aide à l'extraction, à la modélisation ou la validation de besoins. Le but de l'annotation est de repérer automatiquement les phrases qui, dans un texte, expriment des actions, et d'y extraire les circonstances de l'action (contrôleur, finalité, contraintes). Pour y parvenir, on a effectué une analyse discursive des moyens d'expression de l'action sur des textes de SBI industriels, en français et en espagnol. Cette analyse, d'inspiration linguistique, s'appuie sur la méthode d'exploration contextuelle et la théorie de la Grammaire Applicative et Cognitive pour proposer 1) une typologie de verbes de spécification de besoins informatiques, 2) une base de marqueurs linguistiques de l'action organisé sous forme des règles d'exploration contextuelle, et 3) l'implémentation informatique de ces règles dans le système d'annotation EXCOM et l'annotation d'un corpus de textes de SBI industriels en français et en espagnol.*

URL où la thèse pourra être téléchargée :

<http://www.lalic.paris4.sorbonne.fr>

Mlle Véronique MORICEAU (moriceau@irit.fr)

Titre : Intégration de données dans un système question-réponse sur le Web.

Mots-clés : système question-réponse, inconsistance de données, intégration de données, génération de langue naturelle, évaluation.

Title : *Data integration in a question-answering system on the Web.*

Keywords : *question-answering system, data inconsistency, data integration, natural language generation, evaluation.*

Thèse de doctorat en Informatique, Université Paul Sabatier – Toulouse III, UFR MIG et IRIT – UMR 5505, sous la direction de M. Patrick Saint-Dizier, DR CNRS. Soutenue le 15/02/2007.

Jury : Mme Régine André-Obrecht (Pr, Université Paul Sabatier et IRIT, présidente), M. Patrick Saint-Dizier (DR., IRIT, directeur), M. Guy Lapalme (Pr, Université de Montréal & RALI, rapporteur), M. Pierre Zweigenbaum, (DR., LIMSI, rapporteur), M. Christian Rétoré (Pr., Université Bordeaux I & LaBRI, examinateur), M. Patrick Séguéla (Synapse Développement, invité)

Résumé : *Aujourd'hui, le Web met à la disposition du grand public un très grand nombre de données et les systèmes de recherche d'information développés ces dernières années sont des outils pratiques pour qui souhaite trouver une réponse à une question sur tout type de domaine. L'interrogation de ces moteurs de recherche se fait sur la base de mots-clés et ceux-ci proposent à l'utilisateur un ensemble de liens vers des pages Web et/ou des extraits de ces pages traitant du thème de la requête.*

En revanche, là où les dictionnaires, encyclopédies et bases de données fournissent une réponse synthétique et cohérente, une analyse rapide des réponses proposées par les moteurs montrent que celles-ci sont très souvent non pertinentes (car elles ne répondent pas à la question posée), incohérentes, etc. C'est donc à l'utilisateur qu'il revient de trier et de rechercher, au sein de ces pages, la réponse à sa question. Cette démarche entraîne une perte de temps considérable, sans pour autant avoir l'assurance de trouver une réponse correcte.

Le problème qui se pose alors est le problème de la pertinence et de la cohérence des réponses proposées à l'utilisateur. Les systèmes question-réponse proposent une alternative à ces problèmes.

Pour cette thèse, nous nous plaçons dans le cadre des systèmes question-réponse coopératifs sur le Web. Nos principaux objectifs sont de modéliser, concevoir et évaluer un système capable, à partir d'une question en langue naturelle, de rechercher les informations pertinentes sur le Web et de générer en langue

naturelle une réponse synthétique, même quand le moteur de recherche sélectionne plusieurs réponses potentielles. Pour cela, il faut intégrer entre autres les attentes des utilisateurs et des techniques de traitement de l'information. Travaillant en domaine ouvert, nous nous sommes plus particulièrement intéressés à deux types de questions : les questions attendant des réponses temporelles (dates) et des réponses numériques.

Notre système a donc pour but :

- *l'intégration d'informations provenant des différentes réponses potentielles à une question en utilisant une base de connaissances et des connaissances directement extraites des pages Web. Ce composant permet notamment de détecter les incohérences de données et de prendre en compte les attentes de l'utilisateur pour produire une réponse appropriée ;*
- *la production en langue naturelle de réponses synthétiques et pertinentes vis-à-vis de l'utilisateur. La coopérativité se situe à plusieurs niveaux : il faut produire des réponses courtes, intelligibles et qui expriment le savoir-faire coopératif mis en oeuvre pour résoudre les incohérences de données.*

Nous proposons également des méthodes d'évaluation adéquates pour jauger le système :

- *d'un point de vue technique en évaluant les performances de chacun des composants du système ;*
- *d'un point de vue cognitif en confrontant les utilisateurs aux réponses produites.*

URL où la thèse pourra être téléchargée :

Contactez l'auteur

M. Yannick PARMENTIER (yannick.parmenier@loria.fr)

Titre : SemTAG, une plate-forme pour le calcul sémantique à partir de grammaires d'arbres adjoints.

Mots-clés : calcul sémantique, grammaires d'arbres adjoints, métagrammaires, sémantique sous-spécifiée, analyse syntaxique profonde.

Title : *SemTAG, a platform for semantic construction with Tree Adjoining Grammars.*

Keywords : *semantic construction, tree adjoining grammars, metagrammars, underspecified semantics, deep parsing*

Thèse de doctorat en Informatique, Université Henri Poincaré & LORIA - UMR 7503, sous la direction de Mme Claire Gardent, DR. Soutenue le 06/04/2007.

Jury : Mme Noëlle Carbonnel (Pr, Université Henri Poincaré & LORIA, présidente), Mme Claire Gardent (DR., LORIA- CNRS, directrice), M. Philippe Blache (DR., LPL-CNRS, rapporteur), Mme Laura Kallmeyer (Projektleiterin, Universität de Tübingen, rapporteur), M. Eric De La Clergerie (CR, INRIA, examinateur), M. Denys Duchier (Pr, Université d'Orléans & LIFO, examinateur).

Résumé : *Dans cette thèse, nous proposons une architecture logicielle (SemTAG) permettant de réaliser un calcul sémantique pour Grammaires d'Arbres Adjoints. Plus précisément, cette architecture fournit un environnement permettant de construire une représentation sémantique sous-spécifiée (Predicate Logic Unplugged (Bos, 1995)) à partir d'une grammaire et d'un énoncé.*

Afin de faciliter la gestion de grammaires de taille réelle, la plate-forme SemTAG intègre un compilateur de méta-grammaires. Le rôle de ce compilateur est de produire semi-automatiquement une grammaire à partir d'une description factorisée. Cette description correspond à (a) une hiérarchie de fragments d'arbres et (b) des combinaisons de ces fragments au moyen d'un langage de contrôle. De plus, chaque arbre ainsi produit peut être équipé d'une interface syntaxe/sémantique à la (Gardent et Kallmeyer, 2003).

La construction sémantique est réalisée à partir du résultat de l'analyse syntaxique. Cette analyse est fournie par un analyseur syntaxique tabulaire généré automatiquement à partir de la grammaire d'entrée au moyen du système DyALog (De La Clergerie, 2005). Cet analyseur produit une forêt de dérivations, qui encode toutes les dérivations, et à partir desquelles les unifications des index sémantiques sont extraites.

Cette plate-forme a été évaluée en termes de couverture sémantique sur la test-suite TSNLP.

URL où la thèse pourra être téléchargée :

<http://www.loria.fr/~parmenti/documents/these-yp.pdf>

M. Mickaël TRAN (miceon@hotmail.com)

Titre : Prolexbase. Un dictionnaire relationnel multilingue de noms propres : conception implantation et gestion en ligne.

Mots-clés : Noms propres, dictionnaire électronique, typologie, ontologie, synonymie, méronymie, accessibilité, XM.

Title : *Prolexbase. An electronic relational multilingual dictionary of proper names: design, implementation and on line management.*

Keywords : *Proper Names, Electronic Dictionary, Typology, Ontology, Synonymy, Meronymy, Accessibility, XML*

Thèse de doctorat en Informatique, Université François Rabelais de Tours, sous la direction de Mme Agata Savary, MC et M. Denis Maurel, Pr Soutenue le 20/10/2006.

Jury : Mme Agata Savary (MC., Université de Tours, co-directrice), M. Denis Maurel (P., Université de Tours, co-directeur), M. Christian Boitet (Pr, Université de Grenoble, rapporteur), M. Jean-Marie Pierrel (Pr, Université de Nancy1, rapporteur), M. Denis Lepsant (Pr, Université de Lille 3, examinateur), M. Pierre Zweigenbaum (Dr, LIMSI-CNRS, examinateur).

Résumé : *Cette thèse présente les différentes étapes de la conception et de l'implémentation d'un dictionnaire électronique relationnel multilingue de noms propres, destiné à des processus automatiques.*

Une étude de différents travaux dans le monde des dictionnaires électroniques et dans le domaine des noms propres en linguistique et en TAL nous a permis de proposer une modélisation des noms propres. Cette modélisation repose sur une architecture en deux parties : une partie qui ne dépend pas des langues et une partie qui dépend de la langue.

La première partie est formée d'un niveau métaconceptuel, regroupant les concepts de type, de supertype et d'existence, et d'un niveau conceptuel, qui comprend le concept de nom propre conceptuel et quatre relations (la synonymie, l'hyponymie, la méronymie et l'accessibilité).

La seconde partie comprend le niveau linguistique (prolexème, alias, dérivés et les relations qui dépendent de la langue) et le niveau des instances (formé de l'ensemble des formes fléchies du prolexème, des alias et des dérivés).

Nous avons implémenté notre modèle sous la forme d'une base de données relationnelle. Une interface de consultation (http://tln.li.univ-tours.fr/tln_prolex/prolex.php) et une interface de travail collaboratif (http://tln.li.univ-tours.fr/tln_prolexbase) ont été créées.

Abstract : *This thesis presents the different stage of the design and the implementation of an electronic relational multilingual dictionary of proper names for automatic process.*

A study of different works in the world of electronic dictionaries and in the domain of proper names in linguistic and in NLP has allowed us to propose a modeling of proper names. This modeling is based on architecture in two parts: a common part for languages and a specific part for a given language.

The first part contains a metaconceptual level, which regroups the concepts of type, supertype and existence, and the conceptual part, which regroups the conceptual proper name and four relations (the synonymy, the hyperonymy, the meronymy and the accessibility).

The second part contains a linguistic level (prolexemes, aliases, derivatives and relations which depend on a language) and an instances level (with the inflected form of the prolexeme, the aliases and the derivatives).

We have implemented our model through a relational database. An interface of consultation and a collaborative work interface have been developed.

URL où la thèse pourra être téléchargée :

http://tln.li.univ-tours.fr/Tln_Livres/Tln_These_Mickael_Tran.html
