



Web Référence

Référencement sur mesure de sites internet

<http://www.web-reference.fr>
18, rue Jean-François Leca
13002 Marseille

contact : Caroline Sarrion : caroline@web-reference.fr - 04 91 91 67 27

>Dossier actualité>> 2 - Une interface de dialogue : Semantia <http://www.semantia.net/>

La position de Semantia, comme pour beaucoup de sociétés émergentes dans le langage naturel, est celle du pragmatisme. Au lieu de proposer une énorme machinerie d'analyse et de connaissances linguistiques, Semantia table plutôt sur une bonne représentation d'un domaine défini précisément pour proposer ses solutions de dialogue en langage naturel.

L'art de bien converser selon Semantia est basé sur le principe d'une scénarisation des dialogues sous forme de scripts. Pour une question posée, cet outil procède à l'extraction de mots-clés, mis en contexte grâce à un raisonnement au plus proche d'une méthode algorithmique : sous telles conditions, tel objet (le mot-clé) est associé à telle action (la réponse). La grande force d'un tel dispositif ? C'est qu'il est le fruit de 15 années d'expérience en messageries télématiques où le but du jeu était de faire durer le plus longtemps possible une " conversation " en créant l'illusion d'un interlocuteur humain. On peut donc imaginer la taille de son capital " réplique du tac au tac ". Mais point d'analyse linguistique dans tout ça, à tout le moins, pas d'analyse syntaxique - la composante lexicale (lexique pourvu de liens synonymiques et hiérarchiques), en revanche, est incontournable : Semantia intègre les **thésaurus** clients.

Et pour quoi faire plus compliqué puisque cela marche bien ainsi ? En effet, on m'a signalé la capacité qu'a cet outil à traiter la **co-référence**. Par exemple, dans un jeu interactif (la pauvreté du dialogue est de mon fait !) :

- j'ai perdu l'arme
- elle est dans le coin
- je ne la vois pas

Et bien, sans se laisser décontenancer par le pronom " la ", ambigu de surcroît avec l'article **homographe** qui passe souvent à la trappe des **mots-vides** éliminés, l'interface continue imperturbablement à fournir des précisions concernant l'arme perdue. Sans analyse linguistique. Tout simplement (?) parce que dans ce contexte de dialogue, il ne peut s'agir que de la fameuse arme.

Faisons tout de même le point sur l'aspect convivial de l'outil qui constitue la valeur ajoutée de ce type d'interface :

- Semantia n'est pas animée ; personnellement cela me gêne moins que les sottises mimiques de Julie.- Semantia n'enregistre pas l'identification de l'utilisateur comme le fait Cybelle qui vous gratifie d'un chaleureux " Nice to see you again " lors des sessions ultérieures.

- J'ai effectué un test un peu tordu pour mesurer les possibilités de gestion de la cohérence d'un dialogue : j'ai posé plusieurs fois la même question successivement (même mots, même syntaxe) : j'ai obtenu la même inébranlable réponse. Je me suis pourtant laissé dire que lors de l'époque héroïque des messageries télématiques, les outils développés par Semantia avaient la possibilité de repérer les questions éludées et de revenir à la charge au cours de la conversation sur ces mêmes questions pour obtenir la réponse souhaitée.- Enfin, une de mes requêtes n'ayant obtenu aucune réponse, j'ai communiqué mon email et j'ai effectivement reçu une réponse, dans les meilleurs délais.

Au demeurant, je vous recommande d'aller tester les outils de Semantia sur le site d'Orphée-Multimédia pour son graphisme très réussi. L'hôte virtuel s'appelle Angelo, et

sert de hot-line automatique pour le roman interactif " La cité des échos ".Prévue pour le début de l'été 2001, l'implantation de Semantia comme hôtesse d'accueil sur le site de la SNCM permettra de mieux juger de son efficacité en recherche d'informations. Restons attentifs aux développements de cette société dont la croissance rapide aura des incidences certaines sur les performances d'analyse linguistiques de ses outils.

Liliane Khouri
Responsable du développement linguistique

Autre article disponible sur le site : 1/ "Vous avez dit "Langage Naturel"?
<http://web-reference.fr/actu/langagenaturel.html>

>Glossaire

Thesaurus

Liste des mots non ambigus qui ont été choisis pour décrire un domaine défini. Ces mots sont appelés "descripteurs". Un thésaurus présente une structure sémantique permettant de regrouper autour du descripteur toute une série de mots satellites grâce à des liens de type hiérarchique, synonymique et de renvoi à des termes approchants ("voir aussi").

Co-référence

Dans un système linguistique, il s'agit de la capacité de faire référence à un même objet par des moyens linguistiques différents. Ex. " Le chat vient de manger. Il n'a plus faim. Cet animal est facile à entretenir". Le chat " il " et " cet animal " font référence au même objet dans ce contexte. Ce phénomène linguistique souligne avec acuité l'inadéquation des méthodes statistiques pour la recherche d'informations : dans l'exemple cité, le mot-clé " chat " n'apparaît qu'une fois, et pourtant d'un point de vue linguistique, c'est bien à trois occurrences de ce référent que nous avons affaire.

Homographe

Qualité de deux mots ayant une orthographe identique pour un sens différent et présentant de ce fait un problème d'identification formelle. Ex. avocat (le fruit, l'homme de loi), la (l'article, le pronom), avions (les avions, nous avions), portes (tu portes, les portes)

Mots vides

Mots ayant peu de signification. La liste de ces mots constituent un "anti-dictionnaire". Ils sont également appelés "mots creux" (Voilà), "mots banals" (Noematics). Notons que justement dans le contexte du langage courant en français ils possèdent un poids très important qui peut changer radicalement le sens d'une phrase. La préposition "de" étant le "mot vide" le plus célèbre, point n'est besoin d'être linguiste pour voir la différence par exemple entre "pomme de terre" et "pomme par terre".