

Human Knowledge ou la linguistique de l'ambiguïté

Pour Human Knowledge, la clé de voûte pour comprendre le sens d'une information est la désambiguïsation... Elle voit dans les applications CtoB une très bonne utilisation de son moteur linguistique MS³.

Le moteur MS³ de Human Knowledge est issu des travaux fondamentaux menés par Kabire FIDAALI, docteur en sciences humaines, et Hermann PRIGNITZ, centralien, sur les mécanismes de déformations du sens – que l'on peut résumer par "linguistique des polysémies et ambiguïtés" – entre 1981 et 2003.

Ces travaux ont abouti à la mise au point du concept de MétaSyntaxe Structurale et Sémantique, concrétisée par la réalisation de deux référentiels-lexiques syntaxiques et sémantiques, l'un général, l'autre contextuel, et d'algorithmes de traitement des données linguistiques. Une première application test de la MétaSyntaxe Structurale et Sémantique, dénommée moteur linguistique MS³, a été finalisée en 1998 pour tester la validité de ce modèle de désambiguïsation de l'information. Cette première ébauche de MS³ fonctionnait sur la base d'un lexique généraliste réalisé sous Fox Pro.

Le fonctionnement satisfaisant de ce premier moteur a permis à ses inventeurs de concrétiser sa version actuelle 100 % Java dans le cadre d'un programme R&D cofinancé par l'ANVAR (Jean-Paul Haton expert) et un Business Angel.

Kabire Fidaali et Hermann Prignitz, travaillent depuis un an sur son usage et sa perspective dans l'entreprise. Le grand thème qui les préoccupe, c'est le CtoB, client to business. Si jusqu'ici le BtoB, BtoC, BtoE tendaient à améliorer la communication de l'entreprise vers la clientèle, le CtoB en revanche traite de la communication du client vers l'entreprise. C'est là où le moteur linguistique MS³ peut faire le plus œuvre utile en automatisant la remontée des informations clients quel qu'en soit le média. Dans ce sens, une masse d'informations par définition non structurées est souvent perdue et doit être traitée manuellement. Des centres de contact aux call centers, des sommes d'argent considérables y ont déjà été investies.

Kabire Fidaali et Hermann Prignitz qui viennent de créer la société Human Knowledge, ont pris contact avec de grands intégrateurs comme SchlumbergerSema et Atos. Ne prétendant pas tout faire, et MS³ pouvant s'intégrer à un grand nombre de produits, ils cherchent en effet des partenaires technologiques. Exemple : ils ne sont pas spécialistes d'IHM. Ils s'engagent en raison de la force et de l'originalité de leur moteur linguistique, à apporter du business à leurs partenaires.

Le paradoxe du TALN

Comme beaucoup d'autres inventeurs dans le domaine linguistique, les fondateurs de Human Knowledge entendent combler un chaînon manquant dans le traitement automatique du langage naturel.

Kabire Fidaali situe le problème dans le "paradoxe" du Traitement Automatique du Langage Naturel (TALN). "Par définition les ordinateurs ne savent travailler que sur des données structurées sur la base de processus logiques où l'incertitude est maîtrisée, alors que le langage naturel véhicule des données non structurées, des procédures linguistiques, où il n'y a qu'incertitude, ambiguïté, polysémie... Pour contourner la difficulté, on apprend à l'ordinateur à travailler dans l'incertitude, - statistiques, logique floue, réseaux de neurones, et simuler

l'imperfection de l'être humain. Les résultats sont médiocres. Cependant, aujourd'hui, tout le monde se contente de la stratégie de contournement.

Notre stratégie est différente. Pour nous, au contraire, la désambiguïisation est la clef de voûte qui permet de comprendre le sens d'une information."

La solution MS³ consiste à proposer un moyen de structurer l'information par le sens, en y incluant notamment un modèle de levée automatique des ambiguïtés et autres polysémies inhérentes à tous les messages véhiculés par le langage naturel.

Ce n'est que lorsque la totalité des ambiguïtés a été levée dans un contexte précis que l'on peut affirmer avoir correctement interprété ce message dans ce contexte et par conséquent l'avoir structuré correctement par le sens. Les résultats fournis par MS³ sont accompagnés de diagnostics qui précisent quelles difficultés ont été rencontrées pendant l'analyse d'un message ; ce diagnostic permet de classer les messages, et résoudre une situation problématique soit par application d'une règle locale, soit par identification d'un défaut d'information. A fortiori, si le diagnostic fourni est " compris ", le déclenchement d'actions automatiques devient envisageable.

Ce que propose Human Knowledge c'est la possibilité de mettre en place ce processus d'automatisation au CtoB. Dans un cas de service de réclamations à propos de factures par exemple, une série de bannettes ou paniers seront modélisés. L'ordinateur triera les messages en fonction des critères suivants, "j'ai compris", "j'ai un doute", "je ne comprends pas". Dans le premier cas, le traitement sera automatique, dans le second, semi-automatique avec des règles locales, dans le troisième, manuel. "On fonctionne comme un entonnoir, note Hermann Prignitz, on répertorie toutes les ambiguïtés possibles et on se met à les éliminer."

Le paramétrage de MS³ pour chaque entreprise demande un travail de 3 à 4 mois de linguiste. Il faut ensuite un accompagnement du produit, réalisé de préférence par un spécialiste du business process collaborant avec un socio-linguiste.

Les applications du moteur linguistique MS³ sont extrêmement variées. Mais elles présentent un intérêt tout particulier à chaque fois que pourra s'exprimer pleinement cette capacité de MS³ à séparer de manière fiable les informations correctement interprétées des autres ; et à fournir pour ces dernières un diagnostic précis et facile à exploiter.

Mireille Boris