

Le travail de recherche s'articule autour de deux axes, l'indexation sémantique et la recherche d'information interactive. Le résultat est concrétisé par l'implémentation d'un moteur de recherche qui permet une navigation intra-documentaire et qui indexe à priori un fond documentaire selon trois axes (temps, espace et phénomène).

L'indexation se compose, d'une phase d'analyse, puis d'une indexation par passage.

Le phase d'analyse se décompose elle-même en deux phases. La première analyse, se base sur la mise en évidence de structure sémantique du langage naturelle dans les corpus pour en extraire une structure symbolique.

Une fois cette structure symbolique construite Géosem passe à une seconde phase d'analyse, l'interprétation. Celle-ci transforme la structure symbolique en entité géographique (type, identifiant, qualitatif) qui est associée à un espace spatial. Le but à ce stade étant que Géosem puisse interpréter les données extraites pour les comparer entre elle.

Une fois que l'on peut interpréter les données, Géosem passe à la phase d'indexation. Ici on indexe des segments de documents. La difficulté qui est posée ici est la méthode à appliquer pour découper les documents. L'invariant étant que les segments soit le plus homogène suivant leurs thématiques. Pour cela Géosem se base sur des modèles linguistiques d'ordre discursif pour décrire les segments on parle alors de cadres de discours. A posteriori cette méthode garantit une bonne homogénéité dans les segments.

Une fois la segmentation effectuée, on passe à la phase d'indexation des segments. Pour cela on recherche pour chaque segment les syntagmes les plus saillants statistiquement de chaque segment par un calcul de type TF.IDF.

La particularité de cette indexation est que l'on s'intéresse à une méthode d'indexation non pas par mots clés mais, comme son nom l'indique par sémantique. Ici on expérimente ce mode d'indexation selon trois axes sémantiques qui sont l'espace, le temps et le Phénomène. Cela améliore l'ontologie des données et évite les problèmes de synonymie.

La qualité de la recherche est améliorée, en prenant mieux en compte l'aspect sémantique d'une requête, les réponses sont plus pertinentes.

Dans le cadre de notre TER, cette partie de l'article nous renseigne sur les méthodes d'indexations, elle s'intègre dans la dimension globale du projet « que pense-t-on de l'agglomération de Montpellier », mais ne s'intègre pas dans notre TER. En effet nous développons des outils visuels et non la phase d'indexation des données.

Le second pivot du travail de recherche est la recherche d'information « interactive ».

Dans un premier temps l'utilisateur formule sa requête (via un formulaire) en discriminant les axes sémantiques qu'il souhaite. Géosem renvoie comme résultat les passages correspondant au plus près aux contraintes.

Dans un second temps Géosem propose à l'utilisateur de modifier les contraintes sur la requête en jouant avec des éléments graphiques. L'utilisateur dispose d'une carte géographique pour modifier l'axe spatial. Il peut agrandir ou rétrécir la boîte géographique correspondant au premier résultat. Il dispose aussi d'une frise chronologique pour modifier l'axe temporel, en modifiant l'intervalle de temps.

Le concept de recherche interactive est innovant et intéressant. Dans le contexte d'un outil de recherche selon des axes temporels et spatiaux. Je n'ai pas connaissance de moteur de recherche qui implémente des composants graphiques pour la recherche d'information.

Ce concept est pertinent. D'un point de vue ergonomique, il permet une navigation plus intuitive et donne une meilleure lisibilité de l'information. Dans notre TER nous souhaiterions pousser le concept en allant plus loin dans la navigation interactive et inclure directement des outils graphiques

comme une frise chronologique et une carte cartographique pour le recherche.

Dans le cadre de notre TER le concept d'indexer des informations suivant leurs sémantiques est très intéressant. Nous souhaitons développer un outil d'aide à la décision dans le contexte de l'aménagement du territoire. En somme comprendre l'environnement territorial pour anticiper et donc prendre de bonne décision.

L'indexation des données par leur sémantique permet de nous rapprocher d'une démarche plus scientifique dans l'analyse. Dans notre TER nous souhaitons croiser des informations (hétérogènes : espace, temps, opinion) pour en abstraire de nouvelles informations et dans leurs finalités permettre d'en comprendre des phénomènes et de prendre des décisions.

Les axes spatial et temporel sont fondamentaux dans l'analyse. Cela nous permet une lecture vertical et horizontale des opinions (que pense t on des hôpitaux de **Montpellier** ? Qu'elle est l'évolution de cette opinion lors de **ses dernières années** ? ). L'axe spatial permet d'étudier la dynamique d'une opinion dans l'espace, croisé avec le temps on peut en observer les mutations.