



Recherche d'Information : calcul de pertinence avec pondération spatiale

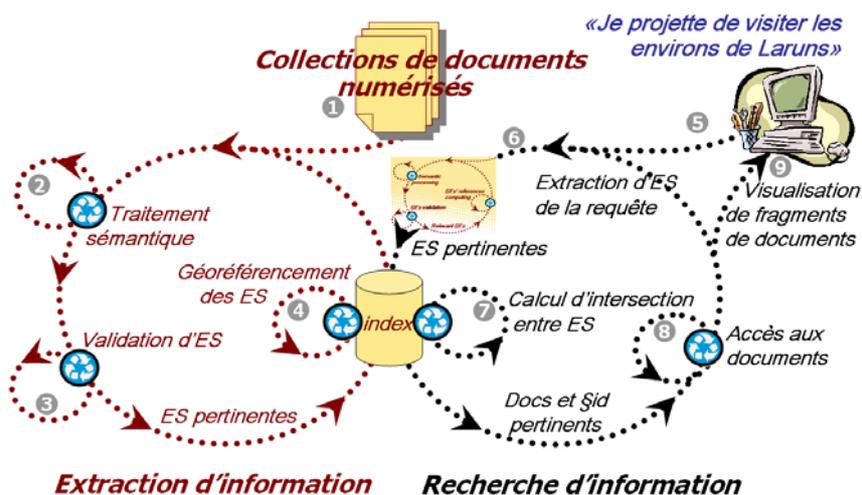
Encadrants du stage : C. Sallaberry (christian.sallaberry@univ-pau.fr), M. Gaio (mauro.gaio@univ-pau.fr), Julien Lesbegueries (julien.lesbegueries@univ-pau.fr)

Personne à contacter : C. Sallaberry (christian.sallaberry@univ-pau.fr), M. Gaio (mauro.gaio@univ-pau.fr)

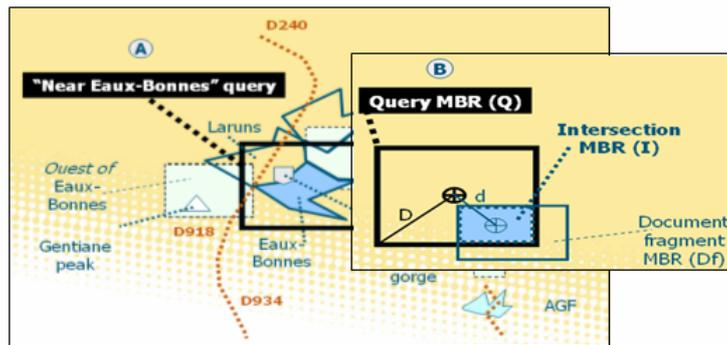
Lieu du stage : LIUPPA PAU

Présentation du sujet

Le projet DESI développé au sein du LIUPPA s'intéresse à la valorisation des corpus documentaires territorialisés. On entend par document territorialisé, tout document qui décrit/raconte un territoire : récits de voyage, contes, cartes postales ... Ce type de document se caractérise par une omniprésence des noms de lieux qui décrivent un territoire particulier. Un premier prototype implémente les phases d'extraction et de recherche d'information spatiales et valide des propositions théoriques originales.

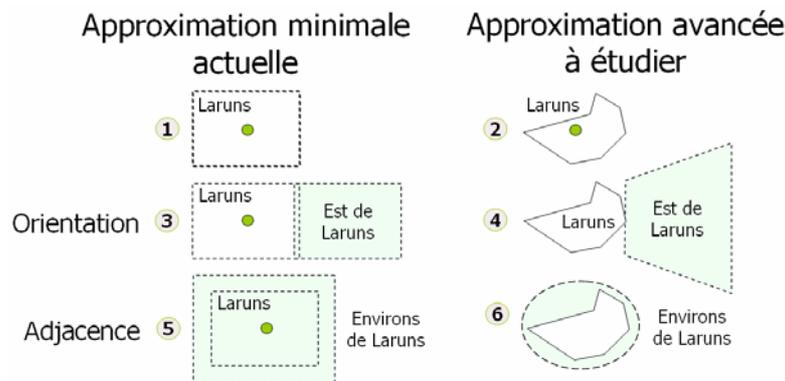


Le stagiaire participera à l'amélioration de la phase 7 de ce processus recherche d'information. Il devra concevoir puis implémenter de nouvelles approches permettant notamment de pondérer plus finement les fragments de document retournés en résultat d'une requête.



Il s'agit, tout d'abord d'évaluer chaque résultat d'une requête (partie A ci-dessus) en utilisant des critères simples de calcul de surfaces d'intersection entre la zone relative à la requête et celle relative au fragment de document retourné par le système de recherche (partie B).

Ensuite, de nouvelles propositions devront tenir compte des représentations spatiales plus fines que les boîtes englobantes présentées ci-dessus. Elles devront, par exemple, intégrer des représentations de ce type dans les algorithmes de pondération :



Il faudra les implanter sur le SIG PostGIS et les expérimenter sur le prototype PIV du projet Desi.

Déroulement du stage

Le stage se déroulera au sein du Liuppa à l'UFR Sciences de l'UPPA. Il se déroulera en 5 grandes étapes :

1. recherche bibliographique relative à la Recherche d'Information Spatiale
2. proposition d'algorithmes de pondération de résultats
3. tests unitaires des algorithmes
4. expérimentations sur le prototype PIV
5. intégration dans le prototype PIV

Mots-clés

Recherche d'Information Spatiale, Calcul de Pertinence, Systèmes d'Information Géographique

Commentaires

Les technologies utilisées seront basées sur les Web Services. Ces derniers feront appels notamment aux fonctions topologiques du SIG PostGIS.

Bibliographie et liens utiles

Sallaberry, C., Gaio, M., Lesbegueries, J., and Loustau, P. *A Semantic Approach for Geospatial Information Extraction from Unstructured Documents*. Chapter for an upcoming book on the Geospatial Web, published by Springer in the Advanced Information and Knowledge Processing Series. To be published in 2007.

Articles relatifs à la Recherche d'Information et au Calcul de Pertinence