

CO-VISUALISATION TEXTE/CARTE POUR L'AIDE À L'ANALYSE D'INFORMATIONS SPATIO-TEMPORELLES EXTRAITES DE RÉCITS

Mots-clés

Analyse visuelle, informations spatio-temporelles textuelles, analyse spatiale, cartographie automatique

Contexte

L'extraction d'informations vise à identifier et produire des informations structurées à partir de documents peu ou pas structurés. Appliquée à des documents textuels, elle consiste à mettre en œuvre des techniques de traitement automatique du langage naturel pour en extraire des informations de catégories prédéfinies (lieux, dates, sentiments, événements, etc.) que l'on souhaite mettre à profit pour une application donnée (recherche d'information, raisonnement, analyses, etc.). La réutilisation de ces informations par des applications suppose leur typage, leur désambiguïsation et leur mise en forme préalables dans un modèle de données structurées (modèle relationnel, RDF, annotations XML, etc.).

Au cours des dernières années, de nombreuses applications d'extraction et de résolution d'entités nommées ont été proposées. Ces applications sont de plus en plus largement réutilisées pour permettre l'extraction d'informations spatio-temporelles dans de grand corpus textuels (articles de presse, blogs, tweets, etc.). Celle-ci constitue la première étape au sein d'applications visant à faciliter la détection d'événements nouveaux, la recherche spatiale d'informations, la cartographie de crise en cas de catastrophes naturelles (Middleton et al, 2014), l'analyse d'opinions sur des politiques d'aménagement du territoire (Kergosien et al, 2015), l'analyse de l'évolution du traitement d'un sujet d'actualité dans les media (Lu et al, 2016), la vérification d'informations issues de réseaux sociaux (Middleton et al. 2016), l'enrichissement automatique de contenus de presse (Gao et al, 2014), etc. Dans toutes ces applications, la visualisation cartographique des informations extraites des textes ou d'informations associées revêt un aspect fondamental pour faciliter l'évaluation des résultats d'extraction, la synthèse des informations extraites, la découverte de connaissances à un plus haut niveau d'abstraction et la communication de ces informations.

Dans le domaine des humanités numériques, de nombreuses sources informations sont disponibles sous la forme de textes. Leur analyse pourrait donc bénéficier des progrès récents en matière d'extraction automatique d'informations à partir de textes et d'analyse visuelle d'informations.

Sujet de stage

Dans ce contexte, l'objectif général du stage est de proposer un prototype d'interface permettant la co-visualisation de textes préalablement annotés et de cartes générées automatiquement à l'aide des informations extraites de ces textes afin de faciliter d'une part, l'évaluation des informations extraites et d'autre part, l'appréhension et l'analyse du contenu des textes.

Nous nous intéresserons ici à un type particulier de textes : des récits traitant de la propagation d'un phénomène sur un territoire. Ces récits peuvent s'accompagner d'indications d'événements survenus dans les lieux traversés (voir section « données disponibles »). Le corpus disponible a fait l'objet d'une annotation préalable. Une étude actuellement en cours a pour objectif d'identifier un ensemble de contraintes cartographiques (détermination d'une échelle de visualisation adaptée selon le type, l'étendue spatiale et le volume des informations à visualiser), d'opérations d'analyse spatiale (détection d'informations aberrantes), de généralisation (agrégation d'informations) et de choix sémiologiques à appliquer afin de produire des cartes efficaces pour chaque type d'analyse visuelle envisagée.

Ce stage aura donc pour objet l'intégration de ces propositions de visualisation cartographique au sein d'une interface Web de co-visualisation carte/texte. Il conviendra donc de proposer des solutions afin d'automatiser l'affichage de cartes aussi conformes que possible aux recommandations de l'étude cartographique préalable. En outre, l'accent devra être mis sur la proposition et la mise en œuvre de solutions d'interaction pertinentes

entre texte et cartes permettant aux utilisateurs de naviguer de façon simple entre représentations cartographiques symbolisées et informations détaillées du texte. En particulier, la gestion de la représentation d'une même information à différents niveaux de détail, au niveau du texte et de la carte et de la coordination de ces différentes représentations devra être étudiée.

Données disponibles

L'interface pourra être testée sur un corpus préalablement annoté, extrait d'un traité d'histoire de Georges Lefebvre, « La Grande Peur de 1789 », retraçant la propagation du phénomène dit de « la Grande Peur » au cours de l'été 1789. Il s'agit d'un mouvement de peur collective attribué en grande partie à une rumeur sur l'existence d'un complot ourdi par l'aristocratie en représailles aux révoltes parisiennes du début du mois de juillet. Les balises identifient les noms de lieux et les dates de passage de la rumeur. Cette annotation pourra être complétée par l'identification des événements permettant de localiser la progression de la rumeur sur le territoire (rapports d'incidents, réactions de panique, agitations, etc.).

Candidatures

Le stage est destiné aux étudiants de niveau master 2 en géomatique ou informatique avec un intérêt pour la géovisualisation. Les compétences requises concernent principalement la modélisation de données (XML, orienté-objet, RDF), l'algorithmique, l'analyse spatiale et le développement d'applications cartographiques pour le Web (Javascript, API cartographiques).

Une candidature doit être accompagnée des documents suivants :

- un CV
- une lettre de motivation
- les relevés de notes de l'année en cours et de l'année précédente

Organisation du stage

Le stage se déroulera dans l'équipe COGIT menant des recherches en géomatique sur la gestion, la diffusion, la représentation et l'utilisation de données géographiques sous forme de référentiels vectorisés et à grande échelle. La durée du stage est de 4 ou 5 mois à partir de février 2017. Le stage se déroulera à l'IGN situé à Saint-Mandé (94).

L'encadrement est effectué par :

Nathalie Abadie (laboratoire COGIT, IGN) nathalie-f.abadie@ign.fr

Sidonie Christophe (laboratoire COGIT, IGN) sidonie.christophe@ign.fr

Bibliographie

Gao, Tong, et al. "NewsViews: an automated pipeline for creating custom geovisualizations for news." *Proceedings of the 32nd annual ACM conference on Human Factors in Computing Systems*. ACM, 2014.

Kergosien, Eric, et al. "Senterritoire pour la détection d'opinions liées à l'aménagement d'un territoire." *Revue Internationale de Géomatique* 25.1 (2015): 11-34.

Lu, Yafeng, et al. "Exploring evolving media discourse through event cueing." *IEEE transactions on visualization and computer graphics* 22.1 (2016): 220-229.

Middleton, Stuart E., Lee Middleton, and Stefano Modafferi. "Real-time crisis mapping of natural disasters using social media." *IEEE Intelligent Systems* 29.2 (2014): 9-17.

Middleton, Stuart E., and Vadims Krivcovs. "Geoparsing and geosemantics for social media: spatio-temporal grounding of content propagating rumours to support trust and veracity analysis during breaking news." *ACM Transactions on Information Systems* 34.3 (2016): 1-27.