

Développement d'outils d'analyse géo-spatiale pour la vente et la distribution : Méthodologie et plateforme.

GAUTIER DARAS¹, BRUNO AGARD², HADRIEN CAMBAZARD³, BERNARD PENZ⁴

¹ FORAC, Ecole Polytechnique de Montréal
Montréal, QC H3C 3A7, Canada
gautier.daras@polymtl.ca

² FORAC, Ecole Polytechnique de Montréal
Montréal, QC H3C 3A7, Canada
bruno.agard@polymtl.ca²

³ Univ. Grenoble Alpes, G-SCOP, CNRS, G-SCOP
F-38000 Grenoble, France
hadrien.cambazard@grenoble-inp.fr

⁴ Univ. Grenoble Alpes, G-SCOP, CNRS, G-SCOP
F-38000 Grenoble, France
bernard.penz@grenoble-inp.fr

Résumé - Dans de gagner du temps au cours du développement d'outils d'analyses spatiales, ce papier introduit une méthodologie de travail et une plateforme de développement. Les cinq principales étapes de la méthodologie sont le rassemblement des données, la création de données spatiales dérivées, le développement du modèle d'analyse, la réalisation de l'interface utilisateur et enfin la validation de l'outil. La plateforme de développement permet une navigation simplifiée au travers de la méthodologie, et permet le travail collaboratif.

Abstract - In the aim of minimizing time wasting and tedious tasks in spatial analysis tools development, this paper introduces a methodology and a framework. The five steps methodology is composed of data gathering, spatial data management, analysis model development, user interface development and tools validation. The framework, based on client-server architecture, allows collaborative work and easy navigation through methodology steps.

Mots clés - SIG, Géographie, ECD, open source, data mining spatial.

Keywords - GIS, Geo visualization, KDD, open source, spatial data mining.

1 INTRODUCTION

Le domaine de la distribution étant de plus en plus concurrentiel, il devient important pour les compagnies d'améliorer leur fonctionnement. Il existe différentes stratégies pour arriver à s'améliorer, chacune d'elle pourra amener les décideurs à prendre des décisions importantes. Pour réaliser ces choix dans les meilleures conditions, les décideurs doivent avoir accès à une compréhension complète et précise des réalités de l'entreprise. En effet, la combinaison de la vision des dirigeants avec des approches normatives permettant de prendre des décisions plus rapidement et efficacement (Spangler, 1991). Ces approches sont souvent basées sur des analyses des données de l'entreprise. Actuellement, de nombreux chercheurs travaillent sur l'extraction de connaissance à partir des données.

Une particularité du commerce de détail est que l'environnement local des points de vente peut avoir une influence directe sur la réussite (Cliquet, Fady, & Basset, 2006). Il est donc pertinent pour ce type d'entreprises de connaître les facteurs d'environnement influençant les chances de succès de chaque point de vente.

L'intégration de l'analyse des données avec les Systèmes d'Information Géographique (SIG) est une solution pour arriver à comprendre les effets de l'environnement.

Étonnamment, les recherches liées à l'utilisation conjointe de ces deux approches sont encore limitées. Cette combinaison doit être faite avec précaution et développée en collaboration avec les futurs utilisateurs pour leur permettre de produire une

connaissance viable et utile. Les modèles d'analyses doivent être robustes et adaptés, et l'information en résultant doit être facilement accessible et compréhensible.

Nos recherches tentent de combiner l'analyse de données géospatiales avec des outils de visualisation, et cela au travers du développement d'un outil d'analyse commerciale.

Dans un premier temps, l'outil devra permettre aux décideurs de comprendre les effets environnementaux. Dans un second temps, il est prévu que l'outil intègre des systèmes d'aide à la prise de décisions.

Tout d'abord, une méthodologie de développement a été définie, et une plateforme pour son application a été mise en place. Le but de cet article est d'introduire cette méthodologie et de présenter cette plateforme. Après un bref état de l'art, la méthodologie sera présentée. La structure de la plateforme, ainsi qu'une implémentation opérationnelle seront ensuite présentées.

2 ÉTAT DE L'ART

Le concept de l'extraction des connaissances à partir des données (ECD) sera d'abord introduit. Ensuite les SIG seront présentés. Enfin, les possibilités d'intégration de ces deux outils seront expliquées.

2.1 ECD

L'exploration de données a été définie comme « la découverte efficace semi-automatisée d'informations non-triviales, implicites, jusque-là inconnues, potentiellement utiles et compréhensibles à partir d'ensembles de données volumineux » (Anand & Buchner, 1998)

Apte (1997), le définit comme un processus par lequel des informations précises et jusqu'alors inconnues sont extraites de grands volumes de données. Ces informations devraient être sous une forme qui puisse être comprise, sollicitée, et utilisée pour améliorer les processus de décision. Il ajoute que les techniques d'exploration de données peuvent être divisées en trois grandes catégories : la modélisation prédictive, la segmentation, et l'extraction de motifs fréquents.

Cependant comme le dit (Cios, Pedrycz, Swiniarski, & Kurgan, 2007), connaître un grand nombre d'algorithmes d'analyse de données n'est pas suffisant pour réussir avec succès un projet de fouille de données.

Ainsi la fouille de données est souvent utilisée dans le processus d'Extraction de Connaissances à partir de Données (ECD). Ce processus a été introduit par (Fayyad, PiatetskyShapiro, & Smyth, 1996), et consiste en neuf étapes principales : la première étape consiste à développer une compréhension du domaine de l'application, mais aussi d'identifier l'objectif du processus de l'ECD du point de vue du client. En effet, la connaissance du domaine est cruciale dans la sélection des données à intégrer. (Shekhar, Zhang, Huang, & Vatsavai, 2003)

Les deuxième, troisième et quatrième étapes consistent en la collecte de données, le nettoyage, le prétraitement, et la réduction. Ces trois phases devraient rendre les données prêtes à passer au travers du processus de fouille de données

Cela conduit aux cinquième, sixième et septième étapes : trouver une catégorie de modèles d'exploration de données qui

correspond à l'objectif du client, trouver les modèles à utiliser dans cette catégorie, et les utiliser pour trouver un résultat pertinent. Les deux dernières étapes concernent l'interprétation des résultats trouvés, et la mise en place d'actions à partir de la connaissance découverte.

2.2 Systèmes d'Information Géographique

Dans un processus de décision complexe, où une grande quantité de données est impliquée, les méthodes de visualisation sont extrêmement utiles (Mendes & Themido, 2004)

Or, les Systèmes d'information géographique (SIG) sont des « systèmes pour la capture, le stockage, l'extraction, l'analyse et l'affichage des données spatiales automatisés » (Clarke, 1990).

(Hess, Rubin, & West, 2004) avancent que, contrairement à des cartes papier, les SIG sont capables de stocker, manipuler et afficher un ensemble beaucoup plus riche d'informations.

(Mendes & Themido, 2004) avance que les SIG permettent en plus d'effectuer des analyses spatiales basiques, et de présenter les résultats sur des supports adaptés.

Les données spatialisées peuvent être représentés sous différents formats au sein de SIG. Par exemple, les polygones peuvent représenter des bâtiments, des lignes grises pourraient être associées à des routes et des points peuvent représenter des stations de bus. Les éléments semblables peuvent être regroupés dans ce qui s'appelle des couches.

Les couches ne contiennent pas seulement les formes des éléments, elles contiennent également toutes les informations accessibles à partir des données de chaque élément. Par exemple, la longueur des routes, le nombre d'habitants des bâtiments, et pour les stations de bus, les numéros de lignes d'autobus.

La plus grande force des SIG est qu'ils permettent aux utilisateurs de trouver les relations spatiales entre les éléments des différentes couches. Par conséquent, il est possible de créer de nouvelles données pertinentes en fonction de nos intérêts de recherche.

Il pourrait être intéressant de connaître la distance moyenne entre deux stations de bus successives d'une ligne particulière. Pour d'autres fins de recherche, il serait plutôt utile de savoir si chaque bâtiment est compris dans une zone autour d'une station de bus. Beaucoup d'autres opérations spatiales pourraient être calculées pour créer des données pertinentes.

Les SIG peuvent également être utilisés comme des outils de géo-visualisation. (MacEachren & Kraak, 2001) avancent que la géo-visualisation intègre des approches du calcul scientifique, de la visualisation de l'information, de l'analyse de l'exploration de données en plus d'autres approches. Il ajoute que l'utilisation conjointe de ces approches devrait permettre une représentation des données facilement réalisable.

Pour Hernandez (2007), la Géo-visualisation aspire à transformer des grandes quantités de données hétérogènes en informations (données interprétées) et par la suite, en connaissances (compréhension dérivée de l'information).

Les techniques de géo-visualisation visent à exploiter les capacités visuelles-cognitives : tels que la reconnaissance des formes, l'ordonnement et l'interprétation des indices visuels.

Par ailleurs, Andrienko, Andrienko, and Gatalsky (2003) avancent que, en comparaison avec les cartes papier, les outils de visualisation sur ordinateur ont principalement deux nouvelles propriétés : l'interactivité et la dynamique.

2.3 Utilisation conjointe de l'ECD et des SIG

L'intégration de l'ECD et des SIG peut se faire de deux façons principales (Le, 2009).

- La fouilles de données spatiales, qui se compose de l'extraction de connaissances implicites et relation spatiale à partir des données spatialisées (Koperski & Han, 1995). Elle peut être constituée, par exemple, de l'identification de relations entre les éléments spatialisés tels que la proximité de services ou de la présence de la densité.

- Géo-visualisation, qui tire parti du fait que les êtres humains apprennent plus facilement et plus efficacement lorsque le support est visuel plutôt que quand il est textuel ou numérique (Lloyd, 1997). Dans un SIG, une abondance de connaissances pourrait être facilement accessible pour l'utilisateur sans pour autant le submerger. La facilité d'accès à l'information pourrait se faire en permettant à l'utilisateur de se concentrer sur des points d'intérêt, ou en lui permettant d'avoir différents formats de visualisation. En effet, (Tuft, 1998) avance que les différents formats de lecture permettent à l'utilisateur d'accès à différents niveaux d'informations, ce qui simplifie la compréhension.

Comme les résultats bruts de la fouille de données sont généralement incompréhensibles pour les décideurs, ceux-ci devraient être affichés dans des interfaces conviviales, telles que les outils de géo-visualisation. Malheureusement, il n'y a que peu de recherches qui tentent de combiner l'exploration de données et la visualisation spatiale (Mendes & Themido, 2004)

Il en résulte qu'aucune méthodologie spécifique ou plateforme de travail n'ont encore été mises en place.

Dans la partie suivante, une version étendue de la méthodologie de (Fayyad et al., 1996), sera introduite. Cette méthodologie intègre les aspects du traitement des données spatialisées et de la géo-visualisation.

3 METHODOLOGIE

Il y a deux principales différences entre le processus de l'ECD indiqué précédemment (Fayyad et al., 1996) et notre méthodologie (figure 1- Daras et al, 2015). Tout d'abord, dans notre processus, les données spatiales doivent être prises en compte. Deuxièmement, une interface utilisateur doit être développée pour permettre aux utilisateurs finaux de comprendre facilement les résultats du processus d'extraction de connaissances.

Notre première phase, consiste en l'apprentissage de la connaissance métier, la collecte et le formatage des données pertinentes.

Après cette phase préliminaire, la seconde étape a pour but de bien prendre en compte les données spatiales

Création des données spatiales dérivées

Comme nous l'avons vu auparavant, les SIG peuvent nous permettre de trouver de nombreuses relations entre les différentes données spatialisées. Toutefois, cette recherche de relations spatiales doit être faite avec réflexion. En effet, si nous prenons l'exemple de l'analyse d'un réseau de transport en commun, imaginons que nous disposons de la position des arrêts d'autobus, ainsi que de la position des passagers. Si nous souhaitons analyser la qualité de la couverture du réseau nous allons chercher à avoir pour chaque passager la distance au plus proche arrêt de bus. Il faudra donc utiliser des fonctions propres aux SIG pour créer cette nouvelle donnée. Cependant, la relation spatiale extraite doit être adaptée au problème que l'on souhaite traiter. Si l'analyse porte sur la qualité de service on peut imaginer chercher à intégrer des informations comme la distance entre stations de bus, ou même le nombre de passager pour une même ligne de bus, etc. Et ce n'est pas fini, quand on connaît la relation spatiale que l'on souhaite avoir, sous quel format est-elle pertinente ? Si c'est une distance, à vol d'oiseau où à pied ? En temps ou en mètre ?

Cet exemple tente d'expliquer pourquoi l'étape de prétraitement du processus d'ECD se transforme à une seule phase de notre méthodologie avec les données spatiales. Cette phase devra être faite avec une connaissance spécifique du domaine d'application.

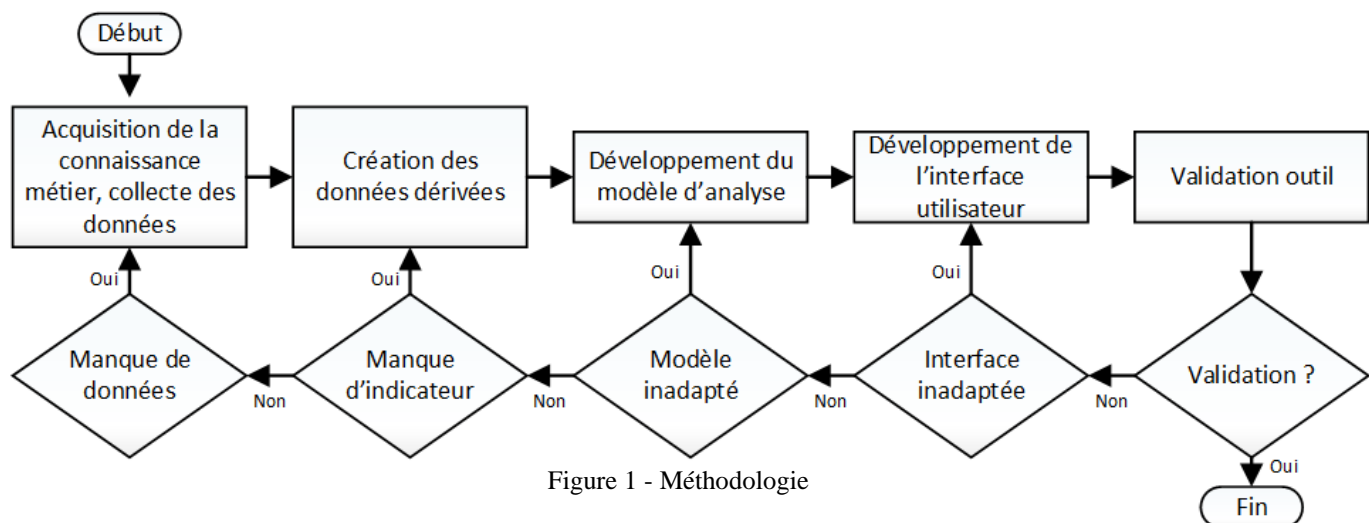


Figure 1 - Méthodologie

Ensuite, vient l'étape du développement du modèle d'analyse, qui consiste à choisir le genre d'information que l'on souhaite extraire, choisir l'algorithme adapté, puis l'appliquer sur les données.

La prochaine étape réside dans le développement d'une interface de visualisation pour l'utilisateur.

Développement de l'interface et validation de l'outil

L'étape de développement de l'interface est ajoutée dans le but de rendre l'information facilement compréhensible pour les décideurs.

Hess et al. (2004) ont avancé que la qualité des outils augmente avec la compétence du décideur. Pour être en mesure d'utiliser cet avantage, la phase de validation des outils est ajoutée pour compléter notre méthodologie. Les deux dernières phases remplacent la neuvième étape initiale.

La phase de développement de l'interface utilisateur doit commencer par la définition des informations qui doivent être accessibles pour les utilisateurs. Les formats de visualisation de ces informations doivent être judicieusement choisis pour optimiser l'expérience utilisateur. Un prototype de l'interface utilisateur doit être implémenté pour s'assurer de sa facilité d'utilisation.

La dernière phase, la validation de l'outil, doit être effectuée avec le client / utilisateur final. Dans cette étape, il sera assuré que les quatre étapes précédentes ont été bien réalisées. L'interface doit être comprise, les options qu'elle offre doivent être utilisées et le modèle devrait fournir des résultats cohérents et utiles. Les chercheurs doivent écouter les observations de tous les utilisateurs et identifier au besoin à laquelle des précédentes phases il faut retourner.

Interface utilisateur inadaptée

L'interface peut être inadaptée pour différentes raisons. Inefficace sur l'ordinateur de l'utilisateur, problème de latence, problème de lisibilité. Quel que soit le problème d'interface, celui-ci renvoie le chercheur à l'étape de son développement.

Manques ou imprécisions dans les données dérivées

Dans le cas de manques ou d'imprécisions dans les données dérivées créées, les chercheurs doivent retourner à la première phase afin d'obtenir de nouvelles données pertinentes, ou à la phase deux, afin de créer de nouveaux indicateurs de valeur ou de relation spatiale.

De nombreux problèmes pourraient retarder ce genre de projets de recherche. Beaucoup de changement de phase se produiront, accompagnés par des transferts de données, ce qui pourrait conduire à des pertes de temps. Pour éviter cela, une plateforme a été conçue et mise en œuvre afin d'appliquer facilement la méthodologie. Ce cadre est présenté dans la section suivante.

4 PLATEFORME

4.1 Éléments à prendre en compte

Des tâches longues et fastidieuses peuvent être rencontrées tout au long de la réalisation de la méthodologie. Les quatre prochaines sections introduisent certains des problèmes identifiés qui peuvent être rencontrés dans ce type de projet. Pour chaque

problème identifié, des outils qui peuvent y remédier seront introduits.

Travailler avec les outils adaptés

Dans chaque projet de recherche, travailler avec des outils adaptés est essentiel. Dans ce projet particulier, nous identifions quatre domaines où les outils adaptés doivent être utilisés : stockage de données, traitement de données, analyse de données, et de visualisation de données. Les quatre domaines concernent les deux types de données : spatiales et non-spatiales.

- Pour le stockage des données, plusieurs bases de données robustes existent sur le marché, mais seulement certaines d'entre elles permettent la gestion des données spatiales. La présence d'une interface d'administration qui permet un traitement rapide est un élément important.

- Pour être en mesure de travailler en profondeur avec des données spatiales, l'utilisation de logiciels SIG puissants et robustes est essentielle.

- Le modèle d'exploration des données doit être réalisé avec des outils rapides et permettant d'utiliser un grand nombre de bibliothèques. Ainsi un grand nombre de modèles facilement paramétrables seront rapidement accessibles.

- L'interface doit être développée avec un outil qui permet de prototyper rapidement. Ainsi, de nombreux formats de visualisation et possibilité d'interaction pourront être implémentés et testés.

Communication entre les outils

Un autre problème est la communication entre les outils. Ils doivent utiliser le même format de données. Les chercheurs ne peuvent pas se permettre les pertes de temps qu'occasionne la conversion du format des données.

- Les interfaces de communication doivent être mises en place entre les différents outils pour automatiser la traduction des données.

Gestion de version et sauvegarde

Un problème fréquemment rencontré est la perte du travail; cela peut provenir de la perte de données, ou du modèle développé qui ne fonctionne plus.

Ne pas être en mesure d'accéder aux versions précédentes du modèle ou à la sauvegarde des données serait un grave problème dans ce cas.

- Les systèmes de gestion de versions ont la possibilité de revenir à une version antérieure d'un document. Ils gardent une trace de toutes les modifications du document, et pour chaque version une sauvegarde est disponible. Un document peut être, par exemple, une sauvegarde de la base de données ou le code source d'un projet.

Design et réalisation de l'interface

Le développement de l'interface utilisateur n'est pas dans le champ de compétence de la plus grande partie des chercheurs. Ainsi, il peut être très chronophage pour un développeur débutant de mettre en place un environnement de géo-visualisation.

- Un Serveur de carte, connecté à la base de données, devrait donner un accès facile à la visualisation de données au travers de technologies web.

Travail collaboratif

Un grand nombre de tâches et de données pourraient être redondantes entre les chercheurs du projet. Par conséquent, une architecture de travail collaboratif qui permet le partage des données ou des codes sources pourrait être un véritable gain de temps.

Dans la section suivante, nous introduisons un cadre basé sur une architecture qui permet le travail collaboratif et l'intégration des outils présentés précédemment.

4.2 Proposition de structure de plateforme

Les problèmes précédemment présentés étant pris en compte, la conception suivante est proposée.

Pour permettre le travail collaboratif, une architecture client-serveur a été choisie. Dans une approche classique serveur lourd-clients le serveur pourrait être surchargé par l'accès aux données et l'analyse géo-spatiale. Dans l'approche du serveur léger / client lourd, les mêmes limitations pourraient se produire sur le côté client (Vatsavai, Shekhar, Burk, & Lime, 2006).

Pour éviter cela, l'architecture que nous proposons est composée de deux types de clients (figure 2): les chercheurs comme des clients lourds et les utilisateurs finaux comme des clients légers.

Ainsi, les chercheurs procéderont à l'élaboration et l'analyse depuis leur propre machine.

L'utilisateur final aura accès à des outils de visualisation via un navigateur internet.

Les clients lourds vont inclure un SIG, des outils d'administration de base de données, et le logiciel d'analyse.

Les serveurs accueilleront les bases de données, les systèmes de

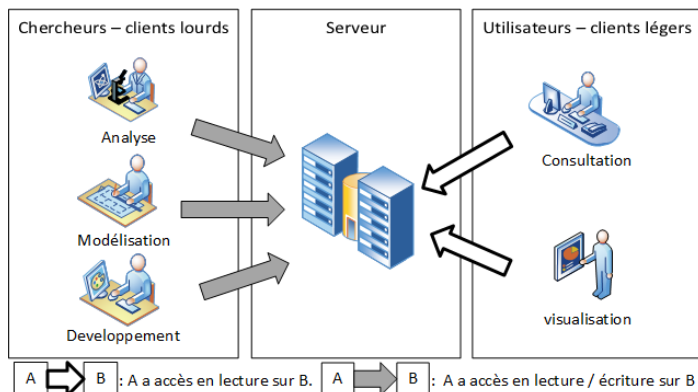


Figure 2 - Architecture

contrôle de version, et un serveur de cartographie Web.

Les utilisateurs finaux (légers) devront juste posséder un navigateur internet conforme sur leurs ordinateurs.

Cette structure de plateforme est l'objet d'une mise en œuvre dans la partie suivante.

4.3 Implémentation de la plateforme

Les besoins technologiques étant désormais présentés, certains logiciels vont être proposés pour réaliser cette plateforme. D'autres logiciels peuvent être mis en place ensemble s'ils ne peuvent pas communiquer entre eux.

L'implémentation de la plateforme a été menée par l'idée de la construire au-dessus de technologies open source.

En plus de permettre une mise en œuvre à relativement faible coût, les technologies open source sont utilisées par un grand nombre de chercheurs permettant la création et le partage de nouveaux logiciels et de connaissances. L'utilisation des technologies open source est la promesse de nouvelles façons d'organiser la science et d'accroître le rythme de découvertes de connaissances selon (Rey, 2009).

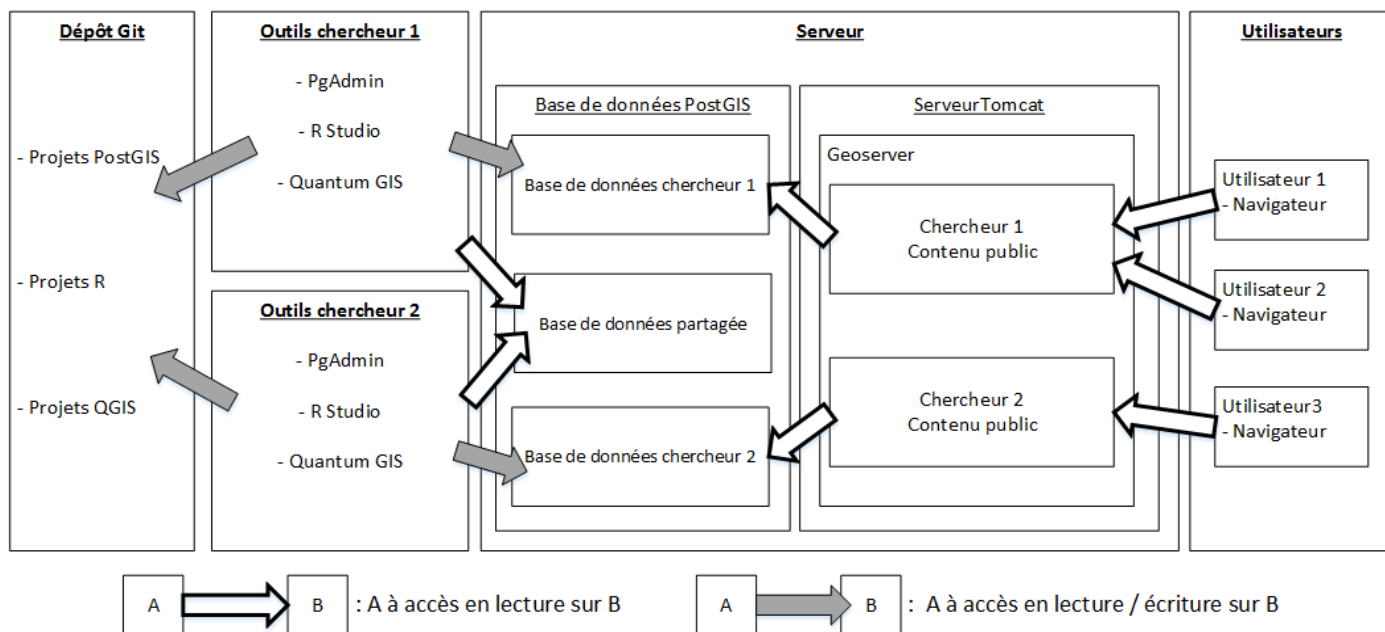


Figure 3 – Plateforme

Dans la section suivante, nous présentons rapidement les logiciels open source qui composent la plateforme, ensuite, le design global sera présenté et expliqué.

La plateforme implémentée avec des logiciels open sources

La plupart des outils présentés ici sont issues de la suite Boundless, anciennement OpenGeo Suite. L'interaction entre ces outils est facilitée car ils font partie de la même collection.

- PostGIS est une extension qui permet aux de bases de données PostgreSQL de prendre en compte les données spatiales.

- GeoServer permet aux utilisateurs de partager et modifier des données géo-spatiales. Il peut distribuer les données de la majeure partie des sources de données spatiales utilisant des normes ouvertes.

- OpenLayers est une bibliothèque JavaScript pour créer des applications de cartographie dans un navigateur.

Les logiciels ne faisant pas partie de la suite boundless sont :

- QGIS est un logiciel de SIG reconnu qui interagit facilement avec une base de données PostGIS.

- R est un langage de programmation reconnu pour le calcul statistique et des graphiques. Il est largement utilisé et possède une grande communauté d'utilisateurs ainsi que de nombreuses bibliothèques d'analyse.

- PgAdmin est une interface d'administration de base de données PostgreSQL qui peut traiter avec base de données PostGIS.

- Git est un système de gestion de version reconnu.

Implémentation fonctionnelle

La mise en œuvre finale est représentée sur la figure 3 (Daras et al, 2015). Il existe quatre blocs principaux, qui ont les caractéristiques suivantes :

Le dépôt git est le lieu où le travail des chercheurs, les algorithmes, et le code source sont placés. Dans cet endroit toute modification est enregistrée, et le code peut être partagé et examiné par des collaborateurs.

Tout chercheur de l'équipe devrait installer les applications

suivantes sur son ordinateur: PgAdmin, R Studio, et QGIS afin de traiter avec l'analyse et la gestion des données.

Le serveur est composé de deux parties principales : le conteneur de base de données PostGreSQL / PostGIS et Geoserver, un serveur cartographique.

- Le conteneur de base de données PostGreSQL / PostGis peut contenir la base de données pour chaque chercheur, et une commune pour les données qui pourraient être accessibles, mais pas modifiables, par chaque membre de l'équipe.

- Geoserver devrait être en mesure de donner aux utilisateurs finaux un accès facile à des données géographiques. Ces données doivent être facilement interprétables dans un navigateur web grâce à l'utilisation d'une bibliothèque de cartographie web comme OpenLayers.

4.4 Discussions / Commentaires

Actuellement, différents logiciels sont combinés afin de naviguer facilement à travers les phases de la méthodologie. Le cadre mis en place permet déjà d'importer facilement différents formats de données spatialisée avec PostGIS. Il est alors possible de traiter ces ensembles de données via PgAdmin pour le nettoyage ou le prétraitement. QGIS peut ensuite être utilisé afin de pratiquer un traitement de données spatiales. Enfin Geoserver peut facilement fournir l'information spatiale au travers de l'outil GeoExplorer ou bien depuis un navigateur Web, en permettant à des bibliothèques de cartographie comme OpenLayers d'accéder à la base de données (figure 4).

Comme les calculs des modèles d'analyse sont effectués sur les ordinateurs des chercheurs, la complexité du modèle devra concorder avec la capacité de calcul des ordinateurs.

Tous les logiciels disponibles pour chaque besoin n'ont pas été testés, ces tests pourraient être utiles pour assurer l'optimisation de l'implémentation de la plateforme.

La partie de la gestion des données est plus développée dans cette plateforme que la partie de visualisation de données. Les recherches futures relatives à développer une interface utilisateur avancée, exigerait des connaissances sur les technologies web (javascript, css, html).

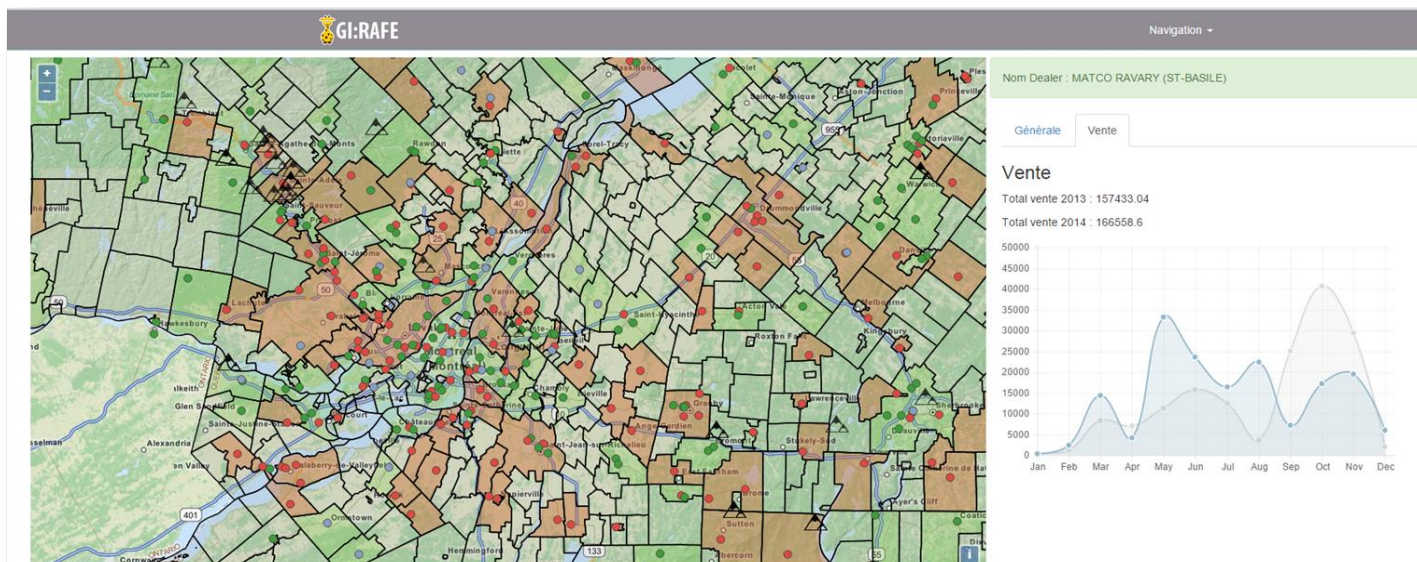


Figure 4 - Interface

Aucune technique d'exploration de données n'a été mise en œuvre pour l'instant, les outils mis en place ont été faits pour permettre l'application de modèles d'exploration de données. Ces modèles devront être judicieusement choisis, et bien adaptés à l'application de domaine.

Enfin, l'outil actuel n'intègre pas de système d'aide à la décision. Dans l'avenir, lorsque les facteurs environnementaux de succès auront été compris, les éléments à optimiser pourront être identifiés. Ensuite, les méthodes de recherche opérationnelle pourraient être adaptées et intégrées dans le système.

5 CONCLUSIONS

Il y avait un manque de méthodologie et d'outils pour l'intégration de la fouille de données spatiales et de la géovisualisation. Cette intégration a d'abord été proposée dans le but de développer des outils d'analyse spatiale.

Pour assurer le bon déroulement de ce développement, un processus d'ECD reconnu a été adapté dans une méthodologie en cinq phases.

Afin de faciliter la navigation à travers les phases de la méthodologie, un ensemble de logiciels open source a été mis en place, résultant en une plateforme de développement.

Des retours sur des applications réelles sont maintenant nécessaires afin d'affiner la méthodologie et le cadre proposé.

6 REFERENCES

- Anand, S. S., & Buchner, A. G. (1998). Decision Support Using Data Mining. *Financial Times*.
- Andrienko, N., Andrienko, G., & Gatalsky, P. (2003). Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages and Computing*, 14(6), 503-541.
- Apte, C. (1997). Data mining: An industrial research perspective. *Ieee Computational Science & Engineering*, 4(2), 6-9.
- Boundless Geo suite website. Retrieved May 15, 2015, from <http://boundlessgeo.com/>
- Cios, K. J., Pedrycz, W., Swiniarski, R. W., & Kurgan, L. (2007). The knowledge discovery process. In Springer (Ed.), *Data Mining A knowledge discovery approach*: Springer.
- Clarke, K. C. (1990). *Analytical and computer cartography*.
- Cliquet, G., Fady, A., & Basset, G. (2006). *Management de la distribution* (2nde ed.).
- Daras, G., Agard, B., Cambazard, H., & Penz, B. (2015). *Development of business spatial analysis tools: methodology and framework*. Paper presented at the 2015 IFAC Symposium on Information Control in Manufacturing – INCOM 2015, Ottawa (Ontario), Canada.
- Fayyad, U., PiatetskyShapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *Ai Magazine*, 17(3), 37-54.
- GeoServer website. Retrieved May 15, 2015, from <http://geoserver.org/>
- Git website. Retrieved May 15, 2015, from <http://git-scm.com/>
- Hernandez, T. (2007). Enhancing retail location decision support: The development and application of geovisualization. *Journal of Retailing and Consumer Services*, 14(4), 249-258.
- Hess, R. L., Rubin, R. S., & West, L. A. (2004). Geographic information systems as a marketing information system technology. *Decision Support Systems*, 38(2), 197-212.
- Koperski, K., & Han, J. (1995). Discovery of spatial association rules in geographic information databases. 47-66.
- Le, T. T. H. (2009). *Perspectives d'intégration entre le datamining et les systèmes d'information géographique*. (M.Sc.), École Polytechnique de Montréal.
- Lloyd. (1997). *Spatial Cognition, Geographic Environments*. Dordrecht: Kluwer Academic Publishers.
- MacEachren, A. M., & Kraak, M.-J. (2001). Research Challenges in Geovisualization. *Cartography and Geographic Information Science*, 28(1), 3-12.
- Mendes, A. B., & Themido, I. H. (2004). Multi-outlet retail site location assessment. *International Transactions in Operational Research*, 11(1), 1-18.
- OpenLayers website. Retrieved May 15, 2015, from <http://openlayers.org/>
- PgAdmin website. Retrieved May 15, 2015, from <http://www.pgadmin.org/>
- PostGIS website. Retrieved May 15, 2015, from <http://postgis.net/>
- QGIS website. Retrieved May 15, 2015, from <http://www.qgis.org/en/site/>
- R Project website. Retrieved May 15, 2015, from <http://www.r-project.org/>
- Shekhar, S., Zhang, P., Huang, Y., & Vatsavai, R. R. (2003). Trends in spatial Data Mining *Data Mining: Next Generation Challenges and Future Directions*: AAAI/MIT Press.
- Spangler, E. W. (1991). The Role of Artificial Intelligence in Understanding the Strategic Decision-Making Process.
- Tufte, R. E. (1998). *Envisioning Information* (6th ed.): Graphic Press.
- Vatsavai, R. R., Shekhar, S., Burk, T. E., & Lime, S. (2006). UMN-MapServer: A high-performance, interoperable, and open source web mapping and geo-spatial analysis system. *Geographic Information Science, Proceedings*, 4197, 400-417.