

Atelier Visualisation d'informations, interaction et fouille de données

Organisateurs :

Hanene Azzag (LIPN, Université de Paris 13),
Pierrick Bruneau (Luxembourg Institute of Science and Technology),
Fabien Picarougne (LINA, Université de Nantes)

PRÉFACE

Désormais bien établi à EGC, l'atelier VIF émane du groupe de travail Visualisation d'Information, Interaction et Fouille de données, fruit de la collaboration entre les associations EGC et AFIHM.

L'atelier se propose de faire le point sur l'actualité en visualisation interactive d'informations, tant du point de vue fondamental que des applications et vise à être un lieu d'échange privilégié entre acteurs scientifiques et opérationnels s'intéressant aux méthodes de visualisation interactive.

À la confluence des communautés EGC et VIS, les méthodes de visualisation interactive et de fouille visuelle des données sont au cœur de cet atelier. Il aura pour vocation de favoriser l'échange sur l'évolution récente des axes de recherche dans ces thématiques, et sur l'application des méthodes de visualisation à des problématiques industrielles. Le traitement de données massives (Big Data) et des flux de données fera l'objet d'une attention particulière.

Hanene AZZAG
LIPN, Université de Paris 13

Pierrick BRUNEAU

Luxembourg Institute of Science and Technology

Fabien PICAROUGNE
LINA, Université de Nantes

Membres du comité de lecture

Le Comité de Lecture est constitué de:

Michaël Aupetit (Qatar Computing Research Institute)	Nicolas Labroche (LI, Université François Rabelais de Tours)
Hanene Azzag (LIPN, Université de Paris 13 Sorbonne)	Guy Mélançon (LABRI, Université de Bordeaux)
Fatma Bouali (LI Tours et Université de Lille2)	Monique Noirhomme (Institut d'Informatique, FUNDP, Namur, Belgique)
Pierrick Bruneau (Luxembourg Institute of Science and Technology)	Benoit Otjacques (Luxembourg Institute of Science and Technology)
Mohammad Ghoniem (Luxembourg Institute of Science and Technology)	Fabien Picarougne (LINA, Université de Nantes)
Fabrice Guillet (LINA, Université de Nantes)	Bruno Pinaud (LABRI, Université de Bordeaux)
Patrik Hitzelberger (Luxembourg Institute of Science and Technology)	Julien Velcin (Université de Lyon 2)
Pascale Kuntz (LINA, Université de Nantes)	Gilles Venturini (LI, Université François Rabelais de Tours)

TABLE DES MATIÈRES

Une architecture flexible de fouille de données visuelle avec R et Javascript <i>Pierrick Bruneau</i>	1
User implicit intention guidance in interactive BI <i>Lingxue Yang, Pierre Morizet-Mahoudeaux, Anne Guénand et Assia Mouloudi</i>	3
Apprentissage et le partage des connaissances humaines avec l'île de mémoire <i>Bin Yang, Chiara Mainardi et Jean-Gabriel Ganascia</i>	5
Protupos : vers un service web de représentations graphiques <i>Sylvie Despres et Jérôme Nobécourt</i>	7
Interprétation visuelle et interactive d'épisodes séquentiels <i>Pierre-Loup Barazzutti, Amélie Cordier et Béatrice Fuchs</i>	9
JASPER: Visualisation orientée pixel de grands graphes <i>Jason Vallet, Guy Melançon et Bruno Pinaud</i>	11
Index des auteurs	13

Une architecture flexible de fouille de données visuelle avec R et Javascript

Pierrick Bruneau

Luxembourg Institute of Science and Technology
5, avenue des Hauts-Fourneaux
L-4362 Esch-sur-Alzette
pierrick.bruneau@list.lu

Résumé. Nous avons récemment mis en ligne un modèle d'application réalisant le lien entre une application web et des services basés sur R : nous proposons une présentation à mi-chemin entre tutoriel et démonstration, permettant de mettre en lumière l'architecture supportant ce modèle, ainsi que quelques recettes et éléments-clés nécessaires à sa bonne mise en oeuvre.

1 Motivation

La plateforme et le langage R sont des outils reconnus pour supporter des tâches d'apprentissage automatique et de fouille de données. Ils sont particulièrement efficaces pour programmer l'exécution de plans d'expériences. De nombreuses fonctions et bibliothèques y sont dédiées à la visualisation de données (e.g., `plot`, `hist`, `lattice`, `grid` (Murrell, 2011), `ggplot2` (Wickham, 2009)), mais les graphiques alors obtenus sont statiques et, au mieux, leur état peut être contrôlé par des lignes de commandes.

Une plus grande interactivité est atteinte avec GGobi (Cook et Swayne, 2007) et Shiny (Shiny, 2014), lesquels autorisent un paramétrage dynamique de graphiques produits par R (e.g., permettant de modifier interactivement les données sous-jacentes à un nuage de points). En revanche, ils sont difficilement extensibles au-delà des scénarios prévus a priori, et ne sont pas prévus pour être intégrés en tant que briques technologiques dans une application plus complexe.

2 Contenu de la présentation

Cette présentation s'intéresse au contexte d'applications web reposant sur des technologies Javascript telles que AngularJS (AngularJS, 2015) et node.js (node.js, 2015). Il est assez facile d'y intégrer des bibliothèques de visualisation interactive très efficaces telles que d3.js (d3.js, 2015) ou Raphaël (Raphaël, 2015). En particulier, la gestion avancée de l'asynchronisme par les moteurs Javascript ouvre la voie au développement d'applications personnalisées combinant visualisation et logique applicative complexe.

Toutefois Javascript n'est pas une plateforme de choix pour le calcul numérique : idéalement ces tâches doivent pouvoir être déléguées de manière asynchrone à un moteur spécia-

lisé tel que R. Nous avons récemment mis en ligne un modèle d'application réalisant ce lien (Rserve-webapp, 2015) : nous proposons une présentation à mi-chemin entre tutoriel et démonstration, permettant de mettre en lumière l'architecture supportant ce modèle, ainsi que quelques recettes et éléments-clés nécessaires à sa bonne mise en oeuvre.

Références

- AngularJS (2015). AngularJS. <https://angularjs.org/>.
- Cook, D. et D. F. Swayne (2007). *Interactive and Dynamic Graphics for Data Analysis : With Examples Using R and GGobi*. Springer.
- d3.js (2015). Data-Driven Documents. <http://d3js.org/>.
- Murrell, P. (2011). *R Graphics*. CRC Press.
- node.js (2015). node.js. <https://nodejs.org/en/>.
- Raphaël (2015). Raphaël. <http://raphaeljs.com/>.
- Rserve-webapp (2015). AngularJS webapp using rserve-js. <https://github.com/pbruneau/Rserve-webapp>.
- Shiny (2014). Shiny. <https://github.com/rstudio/shiny>.
- Wickham, H. (2009). *ggplot2 : elegant graphics for data analysis*. Springer.

User implicit intention guidance in interactive BI

Lingxue Yang^{*,***}, Pierre Morizet-Mahoudeaux^{*}
Anne Guénand^{**}, Assia Mouloudi^{***}

Sorbonne universités, Université de technologie de Compiègne
^{*}UMR CNRS 7253, Heudiasyc, CS 60 319, 60 203 Compiègne cedex
lingxue.yang@utc.fr, pierre.morizet@utc.fr

^{**}Laboratoire COSTECH, CS 60 319, 60 203 Compiègne cedex
anne.guenand@utc.fr

^{***}SAP, 35, rue d'Alsace, 75098 Levallois
assia.mouloudi@sap.com

1 Introduction

Business intelligence (BI) is moving to a new era. With the evolution of user's needs, companies have integrated search capabilities in their BI softwares, which is revolutionizing traditional BI (Becht et de Grauw, 2010). Search-Based BI can provide a variety of answers while the user interacts with the application, mostly by using keywords. The objective is to enable user friendly query languages and answers to facilitate decision making. However, there exists two main problems in this process :

- i) it is hard for users to formulate queries precisely (Ruotsalo et al., 2014). They know their search intention, but they do not know how to express it, and search engines have issues with tasks that go beyond simple keyword queries (Ruotsalo et al., 2013),
- ii) users are not fully aware of which, from amongst the BI processes available, is the most appropriate to resolve their current problem.

Although only the explicit part of the user's intention is input in the BI application, it may infer a part of the hidden or implicit intention to provide the user with some enriched feedback (display, representation, visualization, etc.). Then, usually, the user will reflect, reconsider, rethink his/her intention, according to the feedback, to take the next action. It is then mandatory that the system draws to the attention of the user, what it has "understood" for further proceeding. The question is then, how to display this information, and provide the users with interaction tools, to increase the levels of confidence and trust of the users in interactive BI system, and to give them the feeling that they are well understood and appropriately guided.

2 Related works

Many research focusing on new interface and mode of interaction in exploratory search have been developed when keyword-based search is no longer sufficient. Exploratory search

model like collaborative tagging serves as navigational cues to facilitate the exploration of information (Fu et al., 2010). Relevance feedback improves user's query and facilitates retrieval of information relevant to user's needs (Ruthven, 2003). Faceted search presents a variety of categories among which the user can choose instead of guessing possible keywords. Ruotsalo et al. (2014) have proposed an interactive intent modeling for directing exploratory search. It is based on two principles : radial layout **visualization** of keywords for search intent and directions, and **adaptation** for balancing user exploration and feedback exploitation. However, Ruotsalo's modeling is too general, proposing lot of keywords, which are not easy to select.

3 Research proposal

Starting from (Ruotsalo et al., 2014) approach, we intend to develop a model for interactive intention guidance in the BI domain. We will include in the model, entities and measures such as in (Ballard et al., 2012), but also context information describing the conditions of use (user's experience, history, emotion, ...). The first step has been to select a representative sample of BI users and BI applications. According to a questionnaire that has been built to identify key points of interactivity and guidance functions in various contexts of BI use, we will build an experimental visualization map for search adaptation and guidance.

Références

- Ballard, C., D. M. Farrell, A. Gupta, C. Mazuela, S. Vohnik, et al. (2012). *Dimensional Modeling : In a Business Intelligence Environment*. IBM Redbooks.
- Becht, C. et M. de Grauw (2010). Search-based bi getting ready for the next wave in business intelligence. Technical report, Capgemini.
- Fu, W.-T., T. Kannampallil, R. Kan, et J. He (2010). Semantic imitation in social tagging. *ACM Transactions on Computer-Human Interaction* 17, 1–37.
- Ruotsalo, T., G. Jacucci, P. Myllymäki, et S. Kaski (2014). Interactive intent modeling: Information discovery beyond search. *Commun. ACM* 58(1), 86–92.
- Ruotsalo, T., J. Peltonen, M. Eugster, D. Glowacka, K. Konyushkova, K. Athukorala, I. Kosunen, A. Reijonen, P. Myllymäki, G. Jacucci, et S. Kaski (2013). Directing exploratory search with interactive intent modeling. In *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management, CIKM '13*, New York, NY, USA, pp. 1759–1764. ACM.
- Ruthven, I., L. M. (2003). A survey on the use of relevance feedback for information access systems. *The Knowledge Engineering Review* 18, 95–145.

Apprentissage et le partage des connaissances humaines avec l'île de mémoire

Bin Yang*, Chiara Mainardi**
Jean-Gabriel Ganascia*,**

*Lip6, UMPC-Sorbonne Universités
prénom.nom@lip6.fr,

**Labex OBVIL, Université Paris-Sorbonne
chiara85.mc@gmail.com,

1 Apprentissage à partir de l'île de mémoire HdT

Un corpus littéraire ou numérique est très souvent difficile à apprendre et comprendre. Avec cet article nous introduisons un système de visualisation basé sur la technique des îles de mémoire (YANG (2015)) afin d'aider les utilisateurs à naviguer, comprendre et mémoriser la connaissance à partir du corpus. Le corpus que nous utilisons pour cette recherche est celui du Projet Haine du Théâtre (HdT)¹, qui s'engage dans l'analyse des querelles théâtrales en Europe du XVIIIe au XIXe siècle. A présent, le corpus est composé de 27 textes en langue française.

La visualisation cartographique (visualisation semblable aux cartes) comme la technique des Îles de Mémoire a de nombreux avantages (Gansner et al. (2013)). Elles proposent de transformer et lier les contenus à des représentations bidimensionnelles, qui utilisent des métaphores cartographiques et des analogies géographiques. De plus, elles peuvent aider l'utilisateur à enrichir leur connaissance du domaine et leur permettre de partager leurs opinions et leur compréhension de la visualisation avec d'autres spécialistes. Avec la représentation visuelle (Fig. 1), les utilisateurs peuvent aisément réexaminer les concepts étudiés, revoir la trace de leur visite et des concepts étudiés (avec les fréquences). Nous avons ajouté une fonction « Review » qui permet de revoir les Îles de Mémoire en se basant sur les concepts étudiés précédemment par l'utilisateur. Les utilisateurs peuvent avoir accès à nos résumés sur les concepts avec le logiciel MEDITE (comparer le texte original avec le concept présent dans le corpus). Ensuite, il leur est possible aussi d'améliorer leur niveau de compréhension et de mémorisation grâce à des conclusions préparées et basées sur les concepts étudiés par l'utilisateur. Cette fonction ressemble au processus de la mémoire humaine - elle construit la mémoire à long terme dans notre cerveau. Ces mémoires à long terme (Key Memory) inspire par ailleurs le brainstorming et la naissance de nouvelles idées.

1. <http://obvil.paris-sorbonne.fr/projets/la-haine-du-theatre>

Apprentissage et partage des connaissances humaines avec île de mémoire

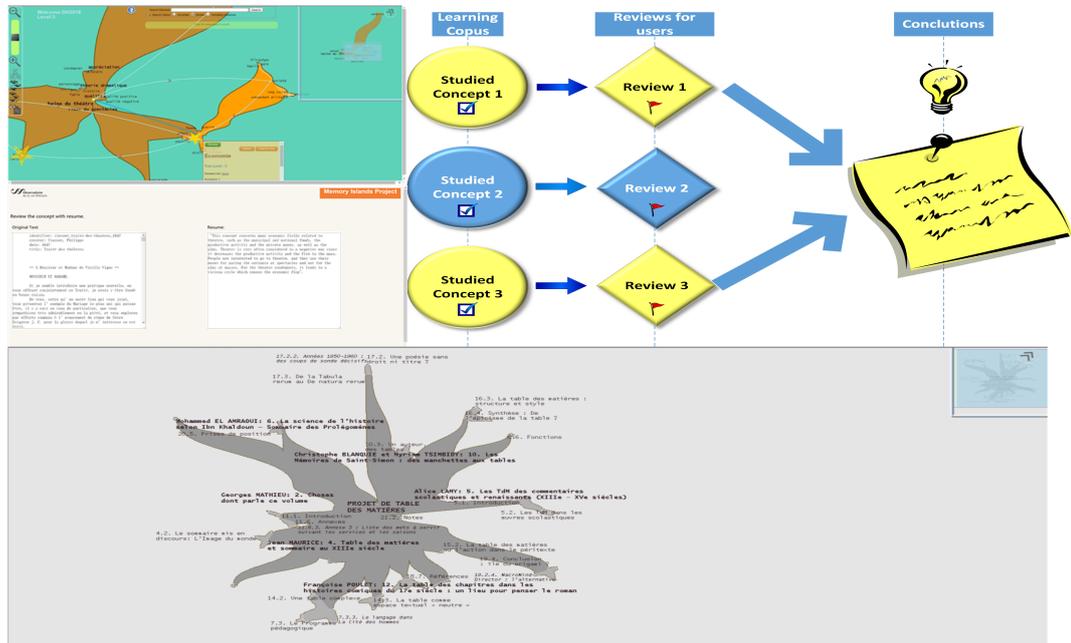


FIG. 1 – L'Île de Mémoire HdT et la fonction "Résumé". Les utilisateurs peuvent voir la conclusion basée sur leur processus d'apprentissage. Dans cette image, l'utilisateur obtient une conclusion des concepts étudiés 1 et 3. En bas : les ToCs (Tables of ToC), les îles de Mémoire peuvent être intégrés dans les Ebook et les corpus en ligne.

2 Applications futures et Conclusions

Nous travaillons également sur une imposante table de matière (Fig. 1) afin de faciliter le partage de la vue d'ensemble d'une collection de tableau de matières. Cette île interactive peut être intégrée très aisément à l'intérieur d'eBooks, ou elle peut également être imprimée dans une version classique du livre. Pour conclure, nous envisageons d'intégrer les techniques NPL et AI, c'est-à-dire la Reconnaissance d'Entités Nommées (NER) ou les analyses de textes (par exemple, la génération automatique de résumés) et, par conséquent, rendre notre technique complètement automatique. Cela permettrait aux spécialistes des humanités numériques de développer leur propre visualisation pour partager ensuite leur connaissance plus facilement.

Références

Gansner, E., Y. Hu, et S. Kobourov (2013). *Handbook of Human Centric Visualization*, Chapter Viewing Abstract Data as Maps. Springer.

YANG, B. (2015). *Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands*. Ph.d thesis, University Pierre and Marie Curie.

Protupos : vers un service web de représentations graphiques

Sylvie Despres*, Jérôme Nobécourt*

*Université Paris 13, Sorbonne Paris Cité, LIMICS, INSERM, (UMRS 1142),
Sorbonne Universités, UPMC Université Paris 06, F-93017, Bobigny, France
prenom.nom@univ-paris13.fr,
<http://www-limics.smbh.univ-paris13.fr/membres/>

1 Introduction

La construction d'ontologies formelles pour un domaine de spécialité est une tâche complexe et coûteuse en temps nécessitant le recours aux experts du domaine. Le noyau des méthodologies de construction d'ontologies comporte quatre étapes fondamentales : acquisition, conceptualisation, formalisation, opérationnalisation. Les deux premières étapes nécessitent notamment l'acquisition et la modélisation des connaissances expertes en les maintenant intelligibles pour les experts. Les experts ne maîtrisent généralement pas les systèmes de dénotation formelle et même si certains le peuvent, ils ne conçoivent pas leur domaine exclusivement au moyen de formules logiques.

2 Vers un service web de représentations graphiques

Les systèmes de visualisation constituent un support alternatif à l'acquisition des connaissances et à leur modélisation. Dans ce contexte, une représentation graphique (RG) produite par l'ontologue ou l'expert, est utilisée comme support à la modélisation. Cette RG peut prendre la forme d'un dessin, d'un schéma, d'une hiérarchie, d'un réseau, d'un graphe, etc. La RG est semi-informelle, elle autorise notamment d'utiliser la langue naturelle pour commenter tout ou partie du modèle. Ces commentaires peuvent prendre la forme d'annotations textuelles, de croquis, de symboles tels que des images ou des icônes, etc. Elles constituent une trace de la construction mentale du modèle qui une fois stockée sert à la capitalisation des connaissances. La RG permet à l'ontologue de formaliser de manière itérative les connaissances représentées tout en laissant suffisamment de liberté à l'expert pour organiser la connaissance qu'il a de son expertise.

Le contenu de la RG évolue au cours de la collaboration entre l'expert et l'ontologue, au fur et à mesure de l'avancement de la construction du modèle. La RG est modifiée par des opérations de transformations (ajout, suppression, modification) traduisant un changement dans le modèle et des annotations. Au cours de la collaboration pour la construction du modèle, différents points de vue peuvent s'exprimer qui conduisent à changer le type de RG utilisé. Par exemple, une représentation hiérarchique sera utilisée pour visualiser l'organisation des

Protupos : vers un service web de représentations graphiques

connaissances, une représentation radiale mettra en évidence le contexte d'utilisation d'une connaissance, un chemin sera utilisé pour expliciter la chaîne de relations liant des connaissances. Les acteurs intervenant autour de cette construction sont à la fois des experts et des ontologues. Par conséquent, la RG permet de confronter leurs idées et de conduire à l'élaboration de RG différentes qui seront conservées. L'obtention d'un consensus n'est pas forcément souhaitable. Au contraire offrir plusieurs « perspectives » pour construire un modèle peut entre autre aider à la construction de ressources modulaires.

Il existe des logiciels de visualisation servant à l'acquisition des connaissances permettant de construire des cartes heuristiques, des cartes conceptuelles, etc. Cependant, les visualisations sont généralement associées à des systèmes propriétaires. Elles servent essentiellement de support aux phases de « brainstorming » mais ne permettent pas une structuration directement exploitable par l'ontologue pour leur formalisation et opérationnalisation. Il existe des plateformes servant à l'opérationnalisation du modèle tels que Protégé pour les logiques de description, CoGui ou Corese pour les graphes conceptuels. Cependant, ces systèmes nécessitent une bonne connaissance des dénotations logiques et la visualisation fournie est totalement dépendante du formalisme choisi. Dans le contexte du « linked data », il existe des solutions pour visualiser les entrepôts de données sous la forme de graphes rdf. Si ces outils sont utiles pour un spécialiste du web sémantique, il est illusoire de penser qu'un expert puisse appréhender la complexité d'une RG sous forme de graphes rdf.

Nous estimons que les techniques et langages qui exploitent les couches sémantiques du « cake du W3C » sont suffisamment matures pour être utilisées dans un service web. Ce service sert alors de médiateur entre une approche non-formelle (lisible par un être humain mais non traitable par une machine) et une approche purement formelle voire même opérationnelle (calculable par une machine, mais difficilement compréhensible par un être humain). Suite à ce constat, nous avons élaboré l'architecture d'un service web intitulé Protupos. Deux prototypes ont été testés sur des cas d'usages pour évaluer la faisabilité et la généralité de cette approche. Dans ce papier, nous présentons l'architecture de Protupos dans le contexte de la construction d'une ontologie dans le domaine de la nutrition.

3 Conclusion

Nous travaillons actuellement à une nouvelle version de Protupos. Les pages sont à base de xhtml et de SVG. Elles supportent les actions de base des réseaux sémantiques en fournissant une syntaxe graphique. Elles précisent pour les primitives conceptuelles utilisées (concept, propriété, annotation, icône, dessins) leurs représentations graphiques. Cela permet, grâce à une architecture simple de client/serveur, d'offrir l'accès à différent mode d'utilisation : commenter une représentation semi-informelle et d'en discuter, ou d'accéder à une partie d'une représentation formelle. Dans ce second cas, il est alors possible d'adopter un point de vue particulier pour susciter un échange entre les acteurs de la construction. Ce type de service web sert également à l'ontologue à appréhender des notions complexes telles que les propriétés calculées. Ces dernières peuvent être traduites en utilisant un formalisme dédié (langages à bases de règles). Il peut également arriver qu'il ne soit pas possible de les opérationnaliser (décidabilité).

Interprétation visuelle et interactive d'épisodes séquentiels

Pierre-Loup Barazzutti * Amélie Cordier ** Béatrice Fuchs ***

* LIRIS CNRS, UMR 5205, Université Lyon 1

** amelie.cordier@liris.cnrs.fr

*** LIRIS CNRS, UMR 5205, Université Lyon 3

beatrice.fuchs@liris.cnrs.fr

Durant le processus d'extraction de connaissances à partir de données (ECD), de nombreuses interactions ont lieu avec un utilisateur humain expert du domaine ou analyste qui y joue un rôle moteur. Pour aider l'analyste dans sa tâche, de nombreux outils permettent toutes sortes de visualisations réalistes des résultats, mais leurs capacités d'interaction restent souvent limitées à des manipulations graphiques. Nous proposons une approche interactive et visuelle de l'ECD supportée par un système de gestion de traces. Nous nous focalisons dans un premier temps sur l'étape d'interprétation.

1 Traces et système à base de traces

Nous nous plaçons dans le cadre de l'étude des *traces modélisées* (Champin et al., 2013) où une trace est issue de l'observation d'un phénomène et est constituée d'une séquence d'éléments observés temporellement situés que nous appelons des *obsels*. Une trace est modélisée c'est-à-dire qu'elle se conforme à un *modèle de trace* décrivant les types d'obsels, leurs attributs et leurs relations et permettant d'interpréter la trace afin de faciliter son exploitation ultérieure. Les traces sont manipulées par des opérations élémentaires appelées *transformations* qui sont de différents types : filtrage d'obsels, fusion de traces, etc. Parmi ces transformations, la *réécriture* est une opération qui crée une nouvelle trace appelée *trace transformée* qui vise à augmenter progressivement le niveau de compréhension et d'abstraction de la trace initiale. La réécriture consiste à construire une nouvelle trace t_2 à partir d'une trace primaire t_1 en remplaçant dans t_2 des motifs, c'est-à-dire des séquences d'obsels, de t_1 par de nouveaux types d'obsels résumant chaque motif et en constituant une interprétation. Un système à base de traces est un système permettant de collecter, de traiter et de visualiser des traces. La réécriture de traces se situe au cœur du dispositif mis en place dans l'interprétation interactive.

Ce travail a été réalisé dans le cadre du projet OFS - Open Food System, programme investissements d'avenir.

2 Extraction de connaissances à partir de traces

Les traces sont étudiées dans le cadre d'un processus d'ECD. Durant le post-traitement, les motifs issus de la fouille sont proposés à l'analyste qui y choisit ceux qu'il estime les plus pertinents compte tenu de ses connaissances du domaine. Le principe de l'interprétation est de produire une trace transformée par réécriture de la trace initiale où les motifs choisis sont remplacés par de nouveaux types d'observés issus de l'interprétation par l'analyste. Pour appréhender méthodiquement l'étape d'interprétation, nous proposons une démarche itérative pour assister l'analyste dans sa tâche, car la fouille produit souvent un nombre important de motifs caractérisés par une forte redondance combinatoire.

Tout d'abord, la liste des motifs est complétée par des indicateurs de qualité permettant de les trier et ainsi mettre en avant les motifs potentiellement les plus intéressants. Les motifs sont alors affichés et l'analyste peut les sélectionner pour visualiser leurs occurrences dans la trace. Lorsqu'un motif est sélectionné, l'analyste peut décider de créer un nouveau type d'obsel pour mémoriser son interprétation dans la trace transformée. À chaque sélection de motif, la liste des motifs est mise à jour dynamiquement en éliminant toutes les occurrences de motif ayant au moins un obsel en commun avec les occurrences du motif sélectionné. Les indicateurs des motifs restants sont alors recalculés et la liste est triée en conséquence. Les motifs ne satisfaisant plus les contraintes sont éliminés, ce qui diminue graduellement le nombre de résultats restants. Ce filtrage automatique facilite ainsi le travail de l'analyste en lui permettant de se focaliser plus rapidement sur d'autres motifs. Lorsque l'analyste a sélectionné tous les motifs et défini les types d'obsels correspondants, il peut alors réitérer le processus d'ECD avec d'autres contraintes sur la même trace ou sur la trace transformée.

Ce scénario d'interprétation interactive a été mis en oeuvre dans le prototype TRANSMUTE qui est présenté de façon plus détaillée dans l'appel à démonstration (Barazzutti et al., 2016).

Références

- Barazzutti, P.-L., A. Cordier, et B. Fuchs (2016). Transmute : un outil interactif pour assister l'extraction de connaissances à partir de traces. In *EGC-2016*, pp. à paraître.
- Champin, P.-A., A. Mille, et Y. Prié (2013). Vers des traces numériques comme objets informatiques de premier niveau : une approche par les traces modélisées. *Intellectica* (59), 171–204.

JASPER: Visualisation orientée pixel de grands graphes

Jason Vallet, Guy Melançon, Bruno Pinaud

Univ. Bordeaux, LaBRI, UMR CNRS 5800, 33405 Talence, France
prenom.nom@labri.fr

Toute sortes de données peuvent être visualisées. Évidemment, certains types de visualisations sont plus adaptés que d'autres selon les cas d'application : ainsi un système de fichiers est souvent dessiné sous la forme d'un arbre, tandis qu'une vue nœuds-liens est préférée pour visualiser un réseau social. Quand le nombre d'éléments à visualiser devient trop important, ces visualisations sont difficiles à utiliser. Les éléments sont alors difficiles à distinguer, rendant l'extraction visuelle d'information d'autant plus complexe.

Lors de nos premiers travaux sur les phénomènes de propagation dans les réseaux sociaux (Vallet et al., 2015), nous nous sommes retrouvés confrontés à ce problème dès que les réseaux étudiés comptaient quelques dizaines de milliers d'éléments. Nous avons en particulier essayé les représentations multi-niveaux qui sont communément rencontrées. Elles permettent une exploration intuitive et peuvent être utilisées sur de très grands graphes, néanmoins, les données de départ doivent être hiérarchisées et tous les éléments ne sont jamais montrés en même temps, empêchant ainsi une représentation globale du graphe. D'autres solutions rencontrées utilisent des courbes de remplissage afin de distribuer de manière efficace et compacte les éléments dans l'espace. Ces techniques, telle que Muelder et Ma (2008), autorisent la représentation d'un nombre d'éléments élevé dans un espace restreint. Malgré tout, l'encodage visuel d'informations sur les nœuds est peu visible de par la taille réduite des éléments et nécessite d'alterner constamment entre une vue globale et une représentation plus détaillée.

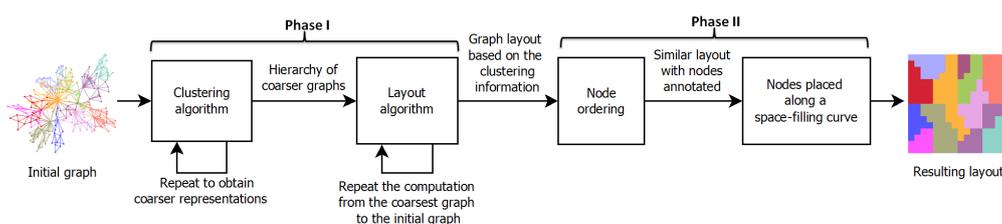


FIG. 1 – Illustration du fonctionnement de JASPER.

Pour répondre à nos besoins, nous avons développé JASPER, qui reprend certaines caractéristiques des représentations orientées pixels, tout en reposant sur l'utilisation d'une courbe de remplissage telle qu'introduite par Keim (2000). La combinaison de ces deux techniques nous permet de répartir les nœuds dans l'espace afin d'améliorer leur visibilité et de quand même prendre en compte les arêtes qui sont souvent absentes des visualisations orientées pixels. En effet, le placement des nœuds sur la courbe est réalisé de façon à préserver la proximité des

JASPER

éléments voisins tels que décrits par la structure topologique du graphe. Nous faisons appel à un algorithme d'agrégation afin de regrouper les éléments par communauté et minimiser la distance spatiale entre voisins lors du dessin (Fig. 1). Notre solution a notamment été utilisée pour visualiser une propagation au sein d'un graphe constitué de plus de cent mille nœuds et d'un million d'arêtes en moins de 10 secondes (voir Fig. 2).

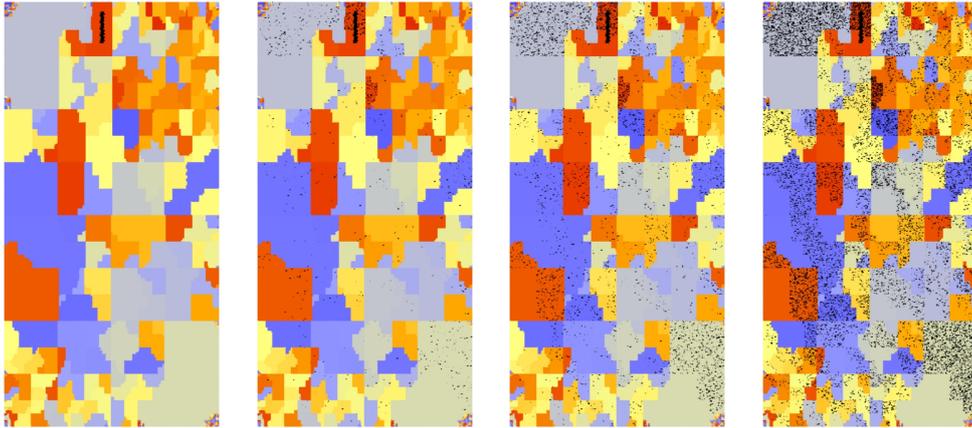


FIG. 2 – Visualisation avec JASPER d'états non consécutifs d'une simulation de propagation dans un réseau (DBLP, 317k nœuds et 1M d'arêtes). Chaque nœud est coloré en noir dès lors qu'il propage l'information à ses voisins.

Références

- Keim, D. A. (2000). Designing pixel-oriented visualization techniques : theory and applications. *IEEE Trans. on Visualization and Computer Graphics* 6(1), 59–78.
- Muelder, C. et K.-L. Ma (2008). Rapid graph layout using space filling curves. *IEEE Trans. on Visualization and Computer Graphics* 14(6), 1301–1308.
- Vallet, J., B. Pinaud, et G. Melançon (2015). Une approche de visualisation analytique pour comparer les modèles de propagation dans les réseaux sociaux. In *EGC 2015*, Volume RNTI-E-28, pp. 365–376.

Index

Amélie, Cordier, 9

Barazzutti, Pierre-Loup, 9

Bruneau, Pierrick, 1

Despres, Sylvie, 7

Ganascia, Jean-Gabriel, 5

Guénand, Anne, 3

Mainardi, Chiara, 5

Melançon, Guy, 11

Morizet-Mahoudeaux, Pierre, 3

Mouloudi, Assia, 3

Nobécourt, Jérôme, 7

Pinaud, Bruno, 11

Vallet, Jason, 11

Yang, Bin, 5

Yang, Lingxue, 3

