

Les bases de données graphes adaptées au Stratifiant.

Discussion de projet

- type de projet : projet FSN (Fiche Stratigraphique Numérique)
- Porteur-r-se-s (si différents des intervenants) : Sogeti
- thématiques de l'atelier concernées : 1 - méthodes et outils d'enregistrement et de traitement des données de terrain ;
- 4- exploration multidimensionnelle de sources croisées pour l'archéologie ;

résumé de présentation

COLLAS, Robin – Scrum Master et Concepteur du projet FSN

Cette présentation a porté sur les réflexions menées par Sogeti sur la réécriture et l'amélioration du Stratifiant (développé par Bruno DESACHY), un outil permettant la détection d'incohérences des données stratigraphiques et la génération du diagramme de Harris. La présentation d'une expérimentation a également été effectuée, permettant de mettre en relief le résultat des travaux réalisés.

Dans le cadre du projet FSN (Fiche Stratigraphique Numérique), un projet d'aide à la saisie numérique des données de terrain, nous avons souhaité intégrer le Stratifiant, une application développée par Bruno DESACHY, un archéologue chercheur auteur d'une thèse sur l'analyse des données stratigraphiques (<https://sitrada.hypotheses.org/outils-daide-au-raisonnement-chronologique-de-terrain>). L'application étant développée en VBA, son intégration dans notre système a posé de nombreux problèmes comme la nécessité d'utiliser un OS Windows, la communication avec une base de données impossible, l'interaction avec le système difficile à mettre en œuvre.

Pour résoudre ces problématiques nous avons donc décidé d'envisager la réécriture complète de cette application dans un langage cross-platform, le Java. Pour ce qui est de l'interaction avec d'autres applicatifs, nous avons aussi repensé l'architecture en appliquant les standards.

Dans le cadre de notre projet R&D nous avons aussi souhaité rendre ce programme dynamique avec une détection d'incohérences réactive, c'est-à-dire dès la saisie de nouvelles données. Nous souhaitons aussi en améliorer les performances pour optimiser les temps de calcul et minimiser les ressources mémoire nécessaires.

Pendant les développements, plusieurs questions se sont donc posées : comment rendre l'algorithme du Stratifiant dynamique ? Quelle est la solution la plus optimale ? Le Stratifiant se basant sur la théorie des graphes, l'utilisation d'une base de données orientée graphes peut-elle simplifier le travail de réécriture ?

Le diaporama de cette présentation est disponible à l'adresse suivante :
<http://sitrada.hypotheses.org/files/2018/07/SITRADA-1.pdf>

discussion

**AH : Amaury Havé ; BD : Bruno Desachy ; JG : Julie Gravier ; PC : Pablo Ciezar ;
RC : Robin Collas ; RD Robert Demaille ; REH : Rachid ElHajaoui ; ALM :
Anne-Laure Morel**

BD : La gestion actuelle des synchronismes avec le système des "bags" [ces ensembles d'US synchrones construits après coup] me paraît poser le problème de la bonne articulation logique des synchronismes (qui sont des relations d'équivalence) et des relations d'ordre (antéro-postériorité). Une unité A synchrone de B est nécessairement antérieure

(sauf incohérence logique) aux unités enregistrées comme antérieures à B : en d'autres termes les déductions logiques de relation d'ordre doivent passer par les synchronismes. La difficulté des outils purement graphe, c'est qu'un diagramme stratigraphique est en fait une superposition de deux graphes (et pas un seul) : celui des relations d'ordre et celui des synchronismes, qui ont des propriétés différentes, mais qui peuvent, et doivent, se combiner pour établir la succession chronologique. La solution retenue pour le Stratifiant est de traiter d'abord les synchronismes, en "fusionnant" les relations d'ordre des unités synchrones (mais en gardant la trace des identifiants de toutes les unités de chaque ensemble synchrone pour pouvoir les afficher sur le graphe).

RD : une des difficultés, c'est aussi de savoir comment gérer la négation/l'interdiction du synchronisme ?

AH : comment se passe la saisie des informations ?

RC : par le biais d'un formulaire sur tablette

REH : Ce mode d'enregistrement par saisie numérique directe qui permettent des retours dynamiques, influe sur l'élaboration des stratégies de fouille, parce qu'il permet de mieux conserver la mémoire de structures qui peuvent disparaître très rapidement sur le terrain, en permettant des retours et corrections rapides de l'enregistrement, alors que le terrain est encore observable. Mais le manque de nuances dans ce dynamisme de la saisie des relations est un danger : on peut être tenté de faire des choix par défaut : si on voit que l'antériorité ne fonctionne pas, alors on rentre une postériorité... Sur le support numérique on rentre « US antérieure à », mais sur la minute de terrain, la fouilleuse/le fouilleur va nuancer son interprétation en inscrivant « antérieure mais... ». Le carnet de fouille revient à la mode : il permet le discours nuancé qui peut ne pas passer dans les systèmes formalisés et numérisés.

RD : il faut toujours laisser la possibilité au fouilleur/à la fouilleuse de rentrer des infos "en vrac", non formalisées, sous forme de rubrique de texte libre notamment, à côté d'éléments descriptifs plus formalisés et analytiques qui permettent d'effectuer des traitements. On peut utiliser des outils classiques (transformés en formulaires sur FSN) ou des bases graphes qui permettent une grande souplesse. Mais peut-être que les graphes ne suffisent pas, peut-être faut-il réfléchir à autre chose ?

BD : y compris avec des outils graphe, il faut poser des règles de cohérence pour éviter d'avoir le graphe ultime (tout est dans tout).

RC : faut-il reléguer les relations stratigraphiques très incertaines à de l'enregistrement type "texte libre", non utilisable pour un traitement ?

BD : Je ne crois pas, car on peut formaliser, et donc traiter, l'incertitude. On peut sortir des diagrammes avec différents niveaux de certitudes ! Mais je pense aussi que pour chaque aspect de l'enregistrement sémantique (c'est à dire non graphique), au-delà de toutes l'information formalisée (les champs à modalités bien définies et donc à choix limité, type "liste déroulante") il est utile et même indispensable d'avoir les rubriques texte libre ("mémo") pour noter les impressions, les nuances, ce qui ne passe pas dans les listes déroulantes. L'intérêt des champs à modalité définies, de l'autre côté, c'est de permettre simplement des traitements statistiques, à commencer par des inventaires précis.

JG : l'enregistrement intégrant des niveaux d'incertitude a aussi tout plein d'avantages pour ce qui est de la gestion des données de fouilles anciennes. Il ne sera plus nécessaire d'avoir à "tuer violemment le père" pour cause d'incompatibilité d'interprétation, mais simplement de préciser *a posteriori* une relation enregistrée au moment de la fouille comme "incertaine".

RC : le stratifiant ne peut analyser que des relations certaines ?

BD : Non, il y a 4 possibilités de saisie : « antérieure », « peut-être antérieure », « synchrone », « peut-être synchrone ». Ces modalités de certitude et d'incertitude permettent un traitement logique : les conflits (ou différences de modalité) entre relations donnent des cas de confirmation, d'infirmité, d'incertitude contradictoire ou de contradiction ; seul le dernier cas est une véritable faute logique, mais les deux derniers exigent l'intervention de l'archéologue, auquel le stratifiant rend alors la main en signalant le problème. Mathématiquement il s'agit d'une forme simplifiée de logique

modale (que l'on peut elle même définir comme une forme discrétisée de logique floue).

RC : Si on laisse la possibilité de choisir le niveau de certitude... ne risque-t-on pas de ne jamais voir de certitude complète ?

BD : Effectivement certains RO ou spécialistes craignent cela, un enregistrement "paresseux", qui refuse de "se mouiller", ou tout devient incertain... Mais c'est un manque de confiance envers les fouilleurs ! Le/la fouilleuse expérimenté-e sait reconnaître les cas où il/elle peut mettre 1 (= niveau de certitude raisonnable), et les cas où l'on arrive pas à être sûr d'une limite.

JG : quand il s'agit d'interpréter des données à l'échelle du site, beaucoup sont souvent très précautionneux. Mais il s'agit de prendre position uniquement sur des relations stratigraphiques, la certitude est plus facile à envisager

RD : je ne suis pas pour chiffrer la valeur d'incertitude. Il vaut mieux mettre selon moi « plus ou moins certain », c'est-à-dire rester dans du qualitatif.

BD : Oui, on peut travailler avec des modalités qualitatives. Mais il serait intéressant d'en avoir plus que seulement 2 (certain / pas certain) comme avec l'actuelle version du stratifiant, pour graduer l'incertitude ; encore les nuances...

RD : on pourrait sortir des graphes très différents...

JG : ... et qui permettraient sans doute de mieux comprendre le site au final

RC : Il est clair que l'outil ne fait pas tout, il faut l'accompagner. Le but de FSN, c'est de faire quelque chose de très réactif et à l'écoute des archéos

REH : le risque c'est qu'il n'y ait pas cet accompagnement, et qu'il ne reste qu'un outil que l'on impose aux équipes. Cela renvoie à la question de la façon dont les agents peuvent s'approprier ces nouveaux outils et les intégrer dans leur pratique. Je pense, d'expérience (en particulier dans le domaine de la photogrammétrie), que c'est une question très importante ; il y a un enjeu archéologique qui est le risque de perte d'information si ça se passe mal.

BD : oui, cela pose aussi la question de l'impact des outils numériques sur les organisations collectives. Des outils informatiques qui posent problème, cela peut être le symptôme d'outils mauvais ou inadaptés à l'organisation de l'équipe, mais aussi de dysfonctionnements dans cette organisation ; dans ce dernier cas, la solution n'est pas forcément un nouvel outil ; c'est d'abord d'analyser et d'améliorer l'organisation.