

# Exploration de données et éinaction

Landy Rajaonarivo, Thomas Alves, Eric Maisel, Pierre De Loor  
LAB-STICC / ENIB

---

**Contexte** : l'objet de ce travail est l'étude de la construction de contextes de lecture d'un champ de connaissances par le parcours de bases de données qui le décrivent.

**Problème** : la question de recherche est la suivante : comment contrôler ces parcours de façon à guider les utilisateurs sans les contraindre et ceci malgré la complexité des bases de données (nombre d'éléments et relations implicites entre éléments).

**Méthode** : la méthode adoptée pour ce travail consiste en une co-construction des contextes de lecture accessibles pour l'utilisateur à travers une métaphore 3d.

**Résultats (attendus)** : une première série d'expérimentations a été menée. Elle montre l'apparition d'une co-construction des contextes de lecture ainsi que l'intérêt de l'utilisation d'une métaphore 3d sur une interface Web classique.

**Conséquences (potentielle)** : la principale conséquence est d'esquisser une nouvelle façon d'explorer des bases de données de grande taille et de permettre la mise en évidence de relations insoupçonnées sans cette approche.

**Lien avec l'éinaction** le paradigme de l'éinaction est utilisé pour guider la conception de l'algorithme de contrôle des parcours de la base de données.

**Mots clés** : éinaction, exploration de bases de données, stigmergie, métaphore 3d

## 1 Contexte

Le problème initial à résoudre concerne la compréhension d'un système complexe (une société par exemple, un éco-système, ...) à partir des informations le décrivant stockées dans une base de données. Ces données sont nombreuses et le plus souvent les relations qu'elles entretiennent les unes avec les autres sont locales, voir implicites. La création des connaissances passe donc par des parcours de la base de données et la constitution de contextes de lecture.

Les moteurs de recherche classiques ne paraissent pas adaptés à cette tâche. Ils ne sont en effet intéressants que lorsque l'utilisateur sait déjà ce qu'il cherche : il peut alors formuler une requête qu'il affinera par la suite de façon à obtenir des résultats plus précis. Mais cette approche n'est plus efficace lorsque le domaine cible lui est inconnu. Une partie de

la tâche à réaliser consiste alors à explorer ce domaine pour en connaître les concepts et les relations entre ces concepts.

L'exploration de bases de données complexes nécessite une algorithmique spécifique pour guider l'utilisateur, en particulier pour éviter qu'il ne se perde dans les informations et assurer qu'il puisse accéder rapidement aux informations les plus pertinentes pour lui. D'autre part ce guidage ne doit pas être trop contraignant de façon à ne pas enfermer a priori l'utilisateur dans un parcours qui ne lui convienne pas. Il faut que l'utilisateur puisse agir de façon autonome tout en étant guidé par ses actions.

Pour cela nous proposons un mécanisme adaptatif mettant en oeuvre une co-construction de parcours de bases de données par un utilisateur et un système d'information.

## 2 Problème

Le problème peut se résumer en quelques questions :

- Comment une exploration de base de données peut elle être énaactive ?
- Une telle approche permet elle d'éviter les problèmes de bulles informatives ?
- Comment montrer que l'aspect énaactif de cette exploration permet une meilleure adaptation de l'exploration pour un utilisateur donné et peut mener à une co-construction de connaissances qu'on peut avoir d'un système complexe ?

## 3 Méthode

Pour répondre aux questions posées ci-dessus nous proposons d'adopter une méthode expérimentale. Elle consiste à (1) proposer un modèle d'exploration de bases de données reposant sur le paradigme de l'énaaction, (2) proposer des propriétés à valider, (3) mettre en oeuvre un prototype et (4) mener des expérimentations et des simulations qui viendront confirmer ou infirmer ces propriétés.

Par la suite nous supposerons disposer de données (les éléments de la base de données à parcourir) et de méta-données (données décrivant les éléments de la base de données). Lorsque ces deux types d'informations seront représentés dans le même espace nous parlerons de (méta-)données.

Le modèle que nous proposons d'étudier repose sur les deux points suivants :

- Toute (méta-)données est non seulement située dans un espace partagé avec l'utilisateur mais également incarnée, c'est à dire perceptible et manipulable par cet utilisateur et dotée d'un comportement qui lui est propre ;
- Le milieu, c'est à dire l'ensemble des (méta-)données, guide l'utilisateur dans sa compréhension d'un système complexe particulier. Le milieu sert donc non seulement à présenter les (méta-)données mais également à contrôler le parcours de l'utilisateur dans la base de données. Le milieu doit donc s'adapter à la fois aux données et à l'utilisateur. D'autre part il doit permettre, par analyse des actions de l'utilisateur d'en inférer les intentions.

L'intérêt d'utiliser un espace familier à l'utilisateur est de permettre de ramener à l'échelle de l'humain un certain nombre de mesures (en les inscrivant dans l'espace) qui sinon resteraient abstraites. C'est le cas par exemple des relations de similitude sémantique entre deux informations. Les incarnations de ces deux informations seront d'autant plus proches qu'elles sont sémantiquement similaires. De la même façon on peut considérer qu'elles seront d'autant plus proches de l'utilisateur (dans l'espace où celui-ci ou son avatar se trouvent) qu'elles correspondent à la recherche de cet utilisateur. Ces relations rendues plus naturelles doivent permettre d'apprécier plus facilement les relations abstraites existant entre les informations et par conséquent de mieux comprendre les systèmes complexes étudiés.

On peut alors considérer une architecture conceptuelle dans laquelle un utilisateur évolue dans un espace 3d qu'il partage avec des incarnations d'informations contenues dans une base de données. Ces incarnations sont dotées d'un comportement qui va les conduire à s'auto-organiser en tenant compte à la fois des autres informations mais également de l'utilisateur. Cet utilisateur, en percevant des motifs, des structures résultant de cette auto-organisation et de leur interprétation va agir dans cet espace pour poursuivre son exploration et provoquer l'apparition de nouvelles structures.

Nous avons mis en œuvre un tel schéma dans le cadre du projet ANR ANTIMOINE. Il s'agissait de pouvoir comprendre certains aspects d'une société en parcourant une base de données inventoriant de l'ordre de 25000 objets du patrimoine Breton. Pour cela l'utilisateur est "immergé" dans un environnement (voir Figure 1) contenant des incarnations de "données" (les images représentant des objets du patrimoine) et des incarnations de "méta-données" (par exemple les portes donnant accès à de nouvelles salles mais également les murs qui délimitent des salles qui contiennent des objets proches les uns des autres).

L'utilisateur se déplace dans l'environnement en tenant compte de la perception qu'il en a. Il peut interagir avec l'incarnation d'une donnée patrimoniale (l'observer, la consulter, la sélectionner) ou avec l'incarnation d'une méta-donnée (par exemple passer d'une salle à une autre en passant par une porte).

L'environnement doit pouvoir réagir à ces différentes actions : par exemple en organisant différemment les objets patrimoniaux contenus dans une salle ou en proposant de nouvelles salles contenant des objets correspondant à des thèmes proposés par l'environnement. Il faut noter que même lorsque l'utilisateur ne fait rien de nouvelles salles peuvent être proposées.

L'adage "La marche pour chemin" est bien mis en œuvre ici puisque les actions de l'utilisateur (la marche) peuvent modifier l'environnement (le chemin) au même titre que l'environnement peut influencer les actions de l'utilisateur (l'existence d'une salle sur un thème donné peut attirer ou au contraire repousser un utilisateur).

L'implémentation repose sur deux points :

- l'incarnation des parcours des bases de données (et non pas seulement des résultats).
- La capacité à pouvoir prédire les intentions de l'utilisateur (et donc à proposer les



Figure 1: Interface 3d : métaphore du "musée virtuel vivant" : l'environnement croît en tenant compte des actions de l'utilisateur et par là même contraindra ces actions dans un futur proche. Le milieu est l'histoire du parcours et le contrôle.

informations les plus pertinentes pour celui-ci).

Cette co-évolution est incarnée par une métaphore 3d dans l'environnement dans lequel l'utilisateur navigue. Cet environnement est constitué d'objets créés et modifiés par le système d'information et avec lesquels l'utilisateur peut interagir.

Le domaine considéré est décrit par une base de données qui contient les connaissances factuelles et par un ensemble de graphes qui représentent des connaissances a priori sur ce domaine (graphes que nous appellerons par la suite structures sémantiques).

- Les sommets de ces graphes sont des concepts dont certains sont utilisés pour annoter les objets de la base de données ,
- Les arcs de ces graphes représentent des relations qui dépendent du domaine considéré. Dans le cas du projet ANTIMOINE il s'agit de relations de subsumption.

Une fonction permet d'associer aux sommets et aux arcs de ces graphes des valeurs qui représentent la pertinence de certains concepts pour l'utilisateur. Cette fonction est dotée d'une dynamique qui lui est propre (propagation de valeurs entre voisins, évaporation) mais qui tient également compte de l'influence de l'utilisateur sur certains concepts à travers les interactions entre celui-ci et des (méta-)données qui lui sont présentées. Ces valeurs qui décrivent la pertinence des concepts pour l'utilisateur interviennent lors de la mise à jour du graphe de scène.

Nous proposons un modèle de calcul qui décrit les relations entre le graphe de scène (représentation de l'environnement 3d présenté à l'utilisateur), les traces d'interaction, la base de données, les structures sémantiques et le profil de l'utilisateur. Ce modèle fait intervenir des processus de Prédiction/Anticipation, de Recommandation, d'Incarnation et de Simulation (Figure 2) :

- un processus de Prédiction / Anticipation dont le rôle est de prédire le futur profil de l'utilisateur en considérant son profil actuel, les traces d'interaction, la base de données ainsi que des connaissances a priori représentées par des structures sémantiques,

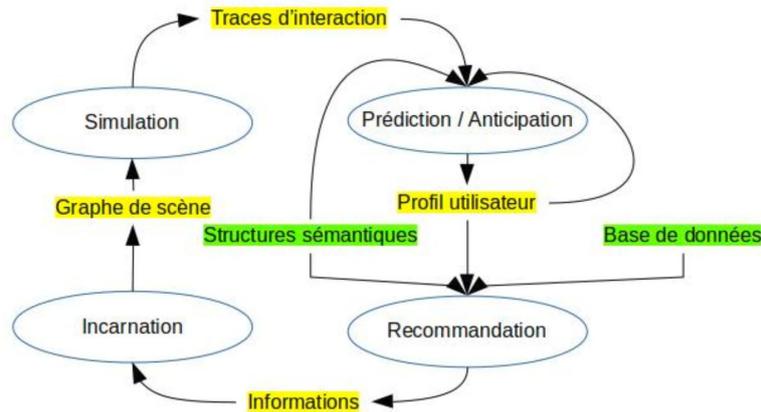


Figure 2: Schéma du processus de recommandation de nouveaux objets à l'utilisateur

- un processus de Recommandation qui détermine quelles sont les informations les plus pertinentes à présenter à l'utilisateur en accord avec le profil utilisateur prédit,
- un processus d'Incarnation qui modifie l'espace perçu par l'utilisateur en y ajoutant non seulement les objets 3d qui correspondent aux incarnations de ces informations les plus pertinentes mais également des représentations 3d des méta-données qui permettent pour certaines d'entre elles de mieux comprendre les relations entre les informations mais également qui donnent des informations quant à la poursuite du parcours de la base de données,
- un processus de Simulation qui conduit à la production d'interactions entre l'utilisateur et l'environnement 3d.

Le processus de prédiction/anticipation met à jour les valeurs de pertinence associées aux sommets et aux arcs des structures sémantiques. Ces valeurs, interprétées comme composants du profil de l'utilisateur, se comportent comme des phéromones : les interactions avec des objets conduisent à des dépôts de phéromones sur les structures sémantiques. Ces phéromones s'évaporent régulièrement.

On peut considérer que le processus de recommandation échantillonne la base de données en accord avec la distribution de ces phéromones. Plus il y a de phéromones plus la probabilité de tirer un concept augmente, et donc ultérieurement les interactions avec les objets qui correspondent à des concepts voisins aux concepts tirés. Une boucle de rétro-action positive émerge de ces interactions en mettant en avant certains groupes de concepts.

Plusieurs scénarii d'utilisation peuvent être envisagés selon les rôles joués par les utilisateurs :

- **Scénario de lecture** : dans ce scénario on suppose que les structures sémantiques ainsi que les annotations qui lient les objets de la base de données et les structures sémantiques sont figées. La tâche de l'utilisateur est de parcourir la base de données pour créer des contextes de lecture et ainsi comprendre le domaine décrit par cette base de données ;

- **Scénario de contribution** : dans ce scénario on suppose que les structures sémantiques sont figées. La tâche de l'utilisateur est de co-construire, avec le système d'information, la base de données, en proposant de nouveaux objets et en les annotant ;
- **Scénario de conceptualisation** : dans ce scénario, les structures sémantiques sont à créer, par co-construction entre l'utilisateur et le système d'information.

Si dans le cadre du projet ANTIMOINE nous nous sommes intéressés essentiellement au premier scénario, nous nous sommes rapidement heurtés au problème de l'annotation automatique d'une grande quantité de documents. Une façon de résoudre ce problème réside dans l'annotation collaborative en particulier en utilisant des folksonomies. Cette approche, que nous avons récemment abordée, permet en partie de traiter le troisième scénario par la création de structures sémantiques à partir des documents et de leurs annotations.

## 4 Résultats (attendus)

Une première série d'expérimentation ont montré d'une part l'intérêt d'une interface 3d sur une interface 2d en particulier en ce qui concerne les capacités à motiver l'exploration et d'autre part les capacités à guider l'utilisateur tout en s'adaptant à ses centres d'intérêt. Ces expérimentations se sont déroulées dans les deux cas sur des temps courts. Il nous reste à étudier la dynamique du couplage sur des temps plus longs et à montrer qu'il y a bien construction de connaissances chez l'utilisateur.

## 5 Lien avec l'énaction

La piste principale adressée par ce document concerne les liens entre l'exploration de données et l'énaction. Ce paradigme peut être un cadre théorique de nombreux domaines ou de nombreux types d'applications. Par exemple, les origines biologiques de l'énaction ont donné naissance à plusieurs travaux en vie artificielle dont l'objectif était de reproduire des principes d'autopoïèse au sein d'un modèle numérique pour approfondir la notion d'autonomie (McMullin 1997, Bourguin et Stewart 2004, McMullin 2004, Ikegami et Suzuki 2008, De Looze et al. 2009). Toujours dans l'objectif de créer une autonomie artificielle, les roboticiens se sont également emparés de ce concept en remplaçant l'architecture cognitive "perception/décision/action" par un système dynamique dont les entrées ne sont que des perturbations d'un mode de fonctionnement autonome (Di Paolo, E., et Iizuka, H. 2008, Di Paolo et al. 2010). Dans le cadre de l'exploration de bases de données, ce n'est pas la notion d'autonomie ou d'autopoïèse qui nous intéresse principalement (même si nous donnons une forme d'autonomie à notre modèles) mais celle de couplage entre environnement et organisme. L'environnement étant une base de données, l'organisme étant un utilisateur. D'après l'énaction, ce couplage est à l'origine de la construction "d'un monde" pour l'organisme. Ses actions modifient son environnement qui par retour va modifier ses perceptions et c'est dans les invariants entre actions et perceptions que se construit "un sens". Dans notre cas, nous parlerons de "compréhension" de la complexité inhérente à de nombreuses données liées entre-elles de façons multiples et pas toujours explicite. Notre proposition peut être rapprochée de celle de (Beust 2010, Ferrari et al 2012) qui convoquent également ce paradigme du même point de vue, mais dans un contexte d'interprétation et de compréhension de documents écrits.

## Références

Pierre Beust (2010) , Ontologies et interprétation des textes sur le Web, 16e Journées de Rochebrune (Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels) "Ontologies", 11 pages.

Bourgine, P., Stewart, J. (2004). Autopoiesis and cognition. *Artificial Life*, 10(3), 327–45. <https://doi.org/10.1162/1064546041255557>

Ferrari S., Maurel F., Pierre Beust P., Mauger S., Holzem M., Baudouin N., Trupin E., Saidali Y., Labiche J., Dionisi D., Pour une recherche d'information et une veille juridique interactives et socio centrées. ENT énaactif veille en droit du transport. *Revue des Sciences et Technologies de l'Information - Série ISI : Ingénierie des Systèmes d'Information*, Lavoisier, 2012, 2, p. 17-40.

Pierre De Loor, Kristen Manac'h, and Alexandra Fronville. Simulation of Abstract Autopoietic Machine. In 10th European Conference on Artificial Life (ECAL), 2009.

Di Paolo, E., Iizuka, H. (2008). How (not) to model autonomous behaviour. *Bio Systems*, 91(2), 409–23. <https://doi.org/10.1016/j.biosystems.2007.05.016>

Di Paolo, E., De Jaegher, H., Rohde, M. (2010). Horizons for the enactive mind: values, social interaction, and play. In *Enaction: Towards a New Paradigm for Cognitive Science.*, 31–87. <https://doi.org/10.1098/rsif.2004.0012>.

Ikegami, T., Suzuki, K. (2008). From a homeostatic to a homeodynamic self. *Bio Systems*, 91(2), 388–400. <https://doi.org/10.1016/j.biosystems.2007.05.014>

McMullin, B. (1997). Computational autopoiesis: The original algorithm. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.15.7359&rep=rep1&type=pdf>

Rajaonarivo L, Courgeon M., Maisel E., De Loor P. (2017) Inline Co-Evolution between Users and Information Presentation for Data Exploration, 2017, IUI2017, 22eme International Conference on Intelligent User Interface, Limassol, Chypre

Rajaonarivo L. (2018) Approche co-évolutive humain-système pour l'exploration de bases de données. Thèse de l'université Bretagne - Pays de Loire (à paraître), Juin 2018.