Cartographie sémantique auto-organisée d'un référentiel de connaissances partagé.

Michel Crampes, Sylvie Ranwez, Jean Villerd

▶ **To cite this version:**

Michel Crampes, Sylvie Ranwez, Jean Villerd. Cartographie sémantique auto-organisée d'un référentiel de connaissances partagé.. 19es Journées Francophones d'Ingénierie des Connaissances (IC 2008), Jun 2008, Nancy, France. pp.161-172. hal-00416696

HAL Id: hal-00416696
https://hal.science/hal-00416696

Submitted on 14 Sep 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cartographie sémantique auto-organisée d'un référentiel de connaissances partagé

Michel Crampes, Sylvie Ranwez, Jean Villerd

Centre de Recherche LGI2P de l'Ecole des Mines d'Alès,
Parc scientifique G. Besse, F – 30 035 Nîmes cedex 1
{prenom.nom}@ema.fr

Résumé : Si les connaissances disponibles en ligne sont massives, leur partage au sein d'une communauté requiert une structuration et des techniques adaptées. En particulier, leur exploitation serait facilitée par une visualisation intuitive et organisée. Nous proposons une méthode, un environnement logiciel, et des heuristiques pour créer des cartes auto-organisées pour un référentiel de connaissances.
L'article présente le contexte réel de notre étude et précise ce qui caractérise un référentiel de connaissances. La problématique de la visualisation et son état de l'art sont ensuite exposés. Notre solution pour visualiser un référentiel de connaissances ainsi que certains outils de navigation sont présentés avant de conclure en ouvrant des perspectives de recherche dans le domaine de la structuration de vues et leur transformation automatiques à l'aide de scénarios.

Mots-clés : Référentiel de connaissances, visualisation de connaissances, cartographie, scénarios, système collectif à base de connaissance.

1 Introduction

Avec la possibilité d'accéder à l'information massivement en ligne, et dans la mesure où la connaissance devient un enjeu majeur des organisations (entreprises, groupes sociaux, centres d'éducation, etc.), deux thèmes de recherche convergent et prennent de l'importance : le Web sémantique, c'est-à-dire la possibilité d'accéder en ligne à de la connaissance indexée et organisée, et les référentiels de connaissances, un moyen de partager la connaissance à l'intérieur d'une structure organisée ou d'un projet porté par un collectif.

Si intellectuellement il est acquis qu'un référentiel de connaissances est profitable pour une organisation, il en existe peu d'exemples réellement opérationnels. Cela peut être dû au caractère récent de la problématique. Mais c'est aussi sans doute dû à des difficultés inhérentes à la matière. Le principe du partage n'est pas en question, au moins dans son fondement culturel. En effet il existe maintenant des environnements de partage massif de la connaissance, comme *Wikipédia*, ou de toute autre *objet* (photos, vidéos, musiques, données personnelles). De plus un grand nombre de langages de description de la connaissance et de technologies Web sont disponibles. Le problème semble donc se situer ailleurs. Dans cet article nous nous intéressons à la visualisation du référentiel qui nous semble être l'un des obstacles majeurs tant à son

usage, qu'à son acceptation. Bien que de nombreux modes de visualisation aient été expérimentés, il ne semble pas qu'une solution réellement pertinente ait émergé. Paradoxalement, des environnements de partage de connaissances comme *Wikipédia* restent sur une représentation traditionnelle des relations entre concepts sous forme d'hyperliens, c'est-à-dire sans réseau sémantique explicite. D'autres environnements de partage comme Youtube expérimentent depuis peu dans leur domaine des formes de projection d'objets par similitude, mais cela reste marginal. Pourtant la représentation de réseaux sous forme de graphes (*graph visualisation*), et la représentation de masses d'information en 2D voire en 3D ont fait l'objet d'intenses recherches pendant de nombreuses années.

Le présent article propose des solutions nouvelles à la visualisation d'un référentiel de connaissances, et, au travers de ces solutions, explore de nouveaux usages. Afin d'illustrer concrètement les idées avancées, nous partirons d'une expérience menée à l'occasion d'un projet de recherche sur la toxicologie nucléaire, ToxNuc-E, qui associe depuis plusieurs années une importante communauté de chercheurs venus de plusieurs disciplines et relevant de différents organismes publics.

A partir de notre expérience sur le projet ToxNuc-E, nous caractérisons un référentiel de connaissances, son contenu et sa mise en œuvre. La visualisation est envisagée comme clef de son acceptation et de son utilisation. Nous proposons une solution pour visualiser les connaissances d'un référentiel ainsi que certains outils de navigation. La conclusion introduit des perspectives de recherche dans le domaine de la structuration de vues et leur transformation automatiques à l'aide de scénarios.

2 Contexte

En 2004, consciente que la maîtrise de la toxicologie nucléaire et de son impact sur l'environnement constitueraient un enjeu majeur des prochaines décennies, une communauté de chercheur s'est regroupée autour de cette thématique. Ce programme transversal, ToxNuc-E[1], réunit actuellement plus de 200 chercheurs et 100 doctorants et post-doctorants en physique, chimie, biologie, etc. appartenant à des organismes différents : CEA, CNRS, INRA, INSERM et MRNT. ToxNuc-E est organisé en 15 sous-projets qui tentent d'identifier les effets toxiques d'éléments chimiques, radioactifs ou non, utilisés dans la recherche et l'industrie nucléaires.

Le partage de l'information est la clé d'une collaboration efficace ; aussi, dès le commencement, une plate-forme collaborative a été mise en place par notre laboratoire. Nous avons été amenés à étudier les aspects sociologiques et techniques du travail collaboratif et avons proposé, entre autres, une analyse et une visualisation des interconnections entre sous-projets, chercheurs et domaines de recherche. Nous avons également proposé une gestion innovante de la base des publications (référentiel de connaissance) avec indexation et recherche d'information basée sur une ontologie du domaine, et statistiques d'utilisation consultables. C'est sur ces différents retours d'expérience que nous basons l'étude qui est présentée dans la suite de cet article.

[1] http://www.toxnuc-e.org/

3 Modèle et usage d'un référentiel de connaissance

3.1 Caractéristiques d'un référentiel de connaissance

L'acception du terme référentiel de connaissances est encore à préciser. Nous retenons celle de (S^t-Amant & Renard, 2004) : '*Le terme de référentiel de connaissances, en anglais* Knowledge Repository*, est surtout utilisé en informatique pour désigner une* « base de données regroupant, dans un réseau sémantique, un ensemble d'informations spécialisées relatives aux activités d'une entreprise ou d'une organisation et qui est consultable par ses membres (Office de la langue française, 2001) *». Par extension, le terme est aussi utilisé pour tout ensemble de connaissances exhaustives et structurées relatives à un domaine ou un projet, dont l'utilisation est partagée par un groupe et souvent supportée par des outils de recherche et d'extraction*'. Selon cette définition, un référentiel de connaissance est caractérisé par (i) la présence d'une base de données, (ii) un réseau sémantique, (iii) un ensemble d'informations spécialisées relatives à une entreprise ou une organisation, et (iv) la possibilité de pouvoir être consulté par ses membres. Le terme peut être étendu à (v) un ensemble de connaissances exhaustives et structurées (vi) des outils d'extraction.

Cette double définition nous apparaît trop précise pour ce qui concerne le caractère exhaustif des connaissances. Ceci nous semble très ambitieux, voire utopique, dans la mesure où ontologiquement tout ensemble de connaissances est évolutif tant dans son contenu que dans sa structure. De plus cette exhaustivité serait contradictoire avec une caractéristique donnée par d'autres auteurs qui considèrent que le référentiel doit pouvoir être en permanence modifié et étendu par les membres de la communauté. C'est par exemple ce que proposent (Martin *et al*., 2005), et des référentiels comme GO en biologie (Ashburner *et al*., 2000).

Cette définition est, à l'inverse, trop imprécise sur plusieurs points, en particulier sur le contenu des connaissances du référentiel, la représentation de ces connaissances et surtout l'utilisation partagée qui en est faite. En ce qui concerne la représentation de la connaissance, il est question d'un réseau sémantique. Ce point nous paraît fondamental car il exclut des bases de connaissances comme *Wikipédia*, et en général des structures de types documents indexés ou de types hyperdocuments. Ce point est mis en relief par (Martin *et al*., 2005). Ce qui caractériserait donc le plus un référentiel de connaissances serait la structuration de la connaissance dans un réseau sémantique (qui reste à préciser), et le caractère partagé entre les membres d'une communauté, qu'elle soit économique, sociale ou culturelle.

3.2 Objectifs et contenu d'un référentiel de connaissances

Mis à part la connaissance qui doit être organisée sous la forme d'un réseau sémantique, les différents auteurs détaillent très peu la nature de ce que contient le référentiel, à moins que tout ce qu'il contient ne soit considéré comme de la connaissance.

Dans l'exemple du référentiel ToxNuc-E, trois objectifs principaux ont émergé en rapport avec l'intention du projet de référentiel et des usages attendus. Le référentiel

documentaire porte sur les articles publiés par les membres du projet. L'objectif est d'avoir une vue d'ensemble de la production scientifique, de son contenu conceptuel, et de pouvoir plus facilement retrouver un article. Le second objectif est de pouvoir situer les individus dans les projets, en regard de leur production scientifique, de rechercher des affinités intellectuelles pour favoriser les collaborations, et de pouvoir identifier quelqu'un qui puisse apporter de l'aide ou un complément d'information sur un point précis. Ce point est d'autant plus important que les individus viennent de différents horizons scientifiques (biologie, chimie, etc.) et qu'il est difficile de construire des passerelles entre eux. Le troisième objectif est à la fois scientifique et humain. Dans la mesure où la toxicologie nucléaire est une discipline scientifique récente, le fait de mettre en relation les individus, les productions scientifiques et le contenu en construction doit permettre de définir une communauté de chercheurs réunie autour d'une ontologie nouvelle, même si celle-ci emprunte ses termes et concepts pour l'essentiel aux autres domaines scientifiques bien établis (biologie, chimie, etc.). Le référentiel doit donc servir à faire émerger une ontologie fédérative à partir de laquelle les individus et la production scientifique doivent pouvoir se situer.

De manière plus ambitieuse, en situant la production scientifique dans le temps en relation avec les concepts de l'ontologie, on peut voir l'évolution conceptuelle du projet dans son ensemble et donc en lire les grandes tendances. Ce dernier point est particulièrement intéressant puisqu'il considère que les outils mis en œuvre dans le référentiel, et en particulier les outils de visualisation, doivent permettre d'effectuer des inférences et des propagations de connaissances.

En résumé, le référentiel ToxNuc-E et sa visualisation doivent permettre d'accéder a minima aux individus, aux projets, aux productions, et à l'ontologie du domaine en construction ainsi qu'aux liens entre ces entités avec la mise en relief d'une composante temporelle. Une fonction implicite de la visualisation est d'être le '*miroir collectif*' de la communauté pour en souligner l'originalité et les synergies possibles.

En conclusion, dans le cadre de ce projet, la connaissance porte essentiellement sur qui travaille ou a travaillé sur quoi, avec qui, dans le cadre de quels sous projets, avec quelle production scientifique, avec quelle évolution dans le temps. Par ailleurs quelles sont les proximités intellectuelles que l'on peut mettre en évidence et donc les suggestions de collaboration, autour de quels concepts ?

3.3 Représentation semi-formelle ou formelle de la connaissance

Si par définition un référentiel de connaissances est structuré par un réseau sémantique, il est nécessaire que celui-ci s'appuie sur une formalisation plus ou moins poussée pour être calculable. (Martin *et al*., 2006) discutent les avantages et les inconvénients d'un langage de représentation formel *versus* un langage semi-formel. Les auteurs concluent que, au moins dans un premier temps, il convient de disposer d'une représentation semi-formelle pour faciliter la participation de chacun à la construction de la connaissance. Mais la construction d'un réseau sémantique est difficile sans formalisation. A l'opposé, l'approche purement ontologique, avec une représentation par exemple en OWL, est très en vogue dans la communauté informatique parce

qu'elle autorise des possibilités de calcul. Mais la participation de membres non informaticiens n'est pas possible sans représentation accessible à tous. Ce caractère de fermeture sociale est accentué par des langages de représentation de connaissances comme RDF.

Les 'Topic maps' et leur langage XTM basé sur XML reposent sur le fait qu'il est possible de décrire de manière très riche un réseau sémantique à partir de relations entre 'topics' (concepts) en laissant en particulier la possibilité d'associer des 'occurrences' aux 'topics', autrement dit des 'instances'. Il est alors possible de marier représentation formelle et contenu informel dans des 'occurrences'. Une occurrence pouvant être humaine (un auteur), la connaissance peut être en partie repérée sans être pour autant formalisée de manière trop poussée. Mais le réseau sémantique décrit en XTM reste inabordable pour les membres de la communauté non spécialistes d'XML.

Les autres représentations à base de *frames* constituent des variantes des formalismes ci-dessus. Il est donc évident que la représentation doit être formalisée pour être calculable, mais que son accès doit être proche d'un langage plus ouvert à un utilisateur non spécialiste. Il doit aussi être global et agréable pour favoriser l'échange. Nous allons voir ci-dessous que nous utiliserons un schéma XML de représentation de graphe, mais que la création d'objets ou de liaisons pourra se faire soit par un tableau Excel, soit directement par l'interface de visualisation, la représentation XML restant cachée au néophyte. Il est possible que, si le besoin l'exige, dans le futur, nous intégrions une feuille de style XSLT pour disposer d'une représentation compatible avec les topic maps (XTM), ou avec OWL.

4 Visualisation d'un référentiel de connaissances

La problématique de la visualisation de l'information s'articule sur trois interrogations essentielles : quelle est la nature de l'information que l'on veut visualiser ? Quel est le volume de cette information ? Quels sont les usages que l'on en attend ? C'est de la réponse à cette dernière question que vont dépendre le mode de visualisation, son intérêt et sa mise en oeuvre. Selon (Schneiderman, 1987) « *The eye is the best way for the brain to understand the world around us* ». Dans sa thèse F. Jalabert reprend l'analyse de (Tufte, 1990) pour avancer que « *l'objectif d'une visualisation est de faciliter l'analyse, la compréhension et le partage de données complexes.* (Arnheim, 1969 ; MacEachrean, 1995; Card *et al*., 1999) *utilisent le terme de 'raisonner visuellement'* » (Jalabert, 2007).

Dans le contexte de la visualisation d'un référentiel de connaissances, nous assignons deux objectifs essentiels à la visualisation : la recherche d'information dans le référentiel et l'échange dans le but de favoriser la constitution d'une communauté.

4.1 Fouille

L'intérêt de la visualisation d'un ensemble d'entités, avec éventuellement un fond de carte physique ou symbolique, porte sur la rapidité d'accès à un objet en particulier à partir d'une représentation mentale, connue par l'utilisateur, de la topologie du

corpus. La visualisation est alors associée à des outils d'exploration en contexte du type « *Overview first*, *zoom and filter*, *then details-on-demand* » (Schneiderman, 1996). La navigation dans un corpus de connaissance ne passe pas nécessairement par une visualisation mais peut être portée par un parcours dans un modèle de connaissances comme une ontologie (Crampes & Ranwez, 2000). Cependant, la visualisation permet de donner l'initiative à l'utilisateur pour rechercher des objets ou explorer des parcours tout en lui donnant un cadre de référence visuel de l'univers exploré.

4.2 Partage et échange

Les données d'un individu ou d'un groupe d'individus sont organisées visuellement afin d'en faciliter le partage et la diffusion. L'utilisation consiste pour chaque membre du groupe à situer ses propres données au sein des données du groupe, et de réutiliser les données des autres en prenant appui sur les siennes.

On retrouve donc un usage de recherche d'information (fouille), qui se situe au niveau individuel et au niveau collectif. Au niveau individuel on trouve la recherche d'un objet d'intérêt soit par son profil, soit par sa proximité avec d'autres objets, soit enfin par ses relations avec les autres objets. Au plan collectif, il est possible d'identifier la communauté par ses objets, leur organisation visuelle selon des critères variables, et les tendances que leur organisation présente.

4.3 Visualisation

Dans la citation de (S^t Amant & Renard, 2004) – c.f. section 3.1, le support par des outils de recherche et d'extraction est présenté comme une composante fréquente des référentiels de connaissances. Les outils de visualisation font de plus en plus partie intégrante de l'outillage.

Synthétiquement, on peut considérer que la visualisation peut prendre quatre formes : une simple succession de listes d'objets selon les relations qui les associent, une représentation locale sous forme d'un petit graphe, une représentation globale sous la forme d'un graphe complexe, ou une projection du réseau avec recherche de proximité sémantique entre les sommets. Les deux premiers cas ne donnent qu'une vision locale du réseau sémantique. Ils ne nous intéressent pas dans la mesure où nous avons montré dans l'usage qu'il était important de situer globalement le référentiel, et donc en avoir une vision globale, pour ensuite raffiner la recherche ou faciliter la manipulation de l'information.

La représentation globale d'un réseau sémantique suppose de faire appel à des outils de visualisation de graphes complexes. Cependant, nous n'en avons pas trouvé d'application réelle. Ceci est sans doute dû à la complexité du réseau qui très vite engendre un graphe illisible. Une variante très pratiquée de la représentation d'un graphe complexe est celle limitée à une simple arborescence. Cette représentation est fréquente car elle est simple à mettre en œuvre et la première qui vient à l'esprit quand on considère uniquement la hiérarchie de classes ou de concepts d'un réseau sémantique. Dans cette veine, *ThinkGraph* s'appuie sur SVG et XTM (*topic maps*) ; *Hypergraph* visualise des arbres hyperboliques. La critique que l'on peut adresser à

cette représentation arborescente est qu'elle met de côté la sémantique du réseau, en particulier elle ne permet pas de distinguer des types, empêchant d'effectuer une fouille guidée par la sémantique. La dernière classe de représentation est la projection sur un plan soit d'un ensemble d'objets reliés (donc un graphe), soit d'un ensemble de vecteurs, chaque vecteur étant associé à un objet. Différentes techniques de projection peuvent être utilisées selon le contexte et la nature des données disponibles, telles que l'Analyse en Composantes Principales (ACP), la projection multidimensionnelle (MDS), la carte de Kohonen (Kohonen, 2001), la projection géométrique comme *FastMap* (Faloustos & Lin, 1995), etc., la liste n'étant pas exhaustive.

Ces différentes approches permettent de faire ressortir des proximités sémantiques qui relèvent soit de proximités vectorielles (similitudes dans un espace à *n* dimensions), soit de liaisons entre objets qui se traduisent par des voisinages en projection dans le plan. Mais elles présentent l'inconvénient de construire des visualisations qui, sauf cas particulier, sont des entrelacs de liaisons difficiles à interpréter, loin des règles d'esthétisme mises en avant par la communauté '*graph drawing'* (Battista *et al*., 1999) comme la limitation du nombre de croisements de liaisons ou l'homogénéité de répartition dans l'espace.

5 Visualisation sémantique d'un référentiel de connaissances

Le problème est donc de pouvoir disposer d'une technique de visualisation qui puisse à la fois respecter au mieux les règles d'esthétisme, les proximités, la lisibilité sémantique, et la capacité à disposer de vues globales et locales, sachant que ces différents objectifs sont souvent contradictoires. De plus, la vue doit pouvoir prendre différentes formes selon l'usage attendu et supporter des milliers d'objets et de liaisons entre ces objets. Nous proposons une méthode, un environnement logiciel, et des heuristiques qui permettent d'aller dans ce sens. L'ensemble tire profit de la sémantique du référentiel de connaissances pour atteindre ces objectifs.

Dans les vues qui suivent, trois classes d'objets du référentiel sont intégrées : les documents (articles scientifiques), les auteurs, et les concepts de l'ontologie du domaine. Par ailleurs, les documents sont reliés à certains concepts de l'ontologie, des liaisons 'est un' existent entre concepts, et des liaisons 'est auteur de' existent entre documents et auteurs. Pour des raisons de lisibilité dans l'article, nous avons limité le nombre d'objets dans les vues présentées à 22 pour les documents, 83 pour les auteurs, et à 60 concepts pour l'ontologie.

Parmi les approches proposées dans la littérature pour favoriser le respect de certaines règles d'esthétisme, nous avons choisi d'implémenter une version de l'algorithme des ressorts (ou Force Directed Placement – FDP, Eades, 1984), car il offre des avantages très intéressants ; en particulier, l'insertion ou la suppression d'un nœud réorganisent le graphe dynamiquement

La figure 1 présente la visualisation du sous ensemble du référentiel de connaissances ToxNuc-E réalisée à l'aide de deux forces : le placement par ressorts et une force de répulsion pour les nœuds concepts. Les effets de transparence que nous détaillerons plus loin gomment en partie la complexité. Mais il est clair qu'une autre

stratégie doit être mise en œuvre pour rendre accessible le graphe tant dans sa compréhension globale que dans ses détails. Nous exposons maintenant cette stratégie alternative basée sur l'exploitation de la sémantique et sur la topologie du graphe.

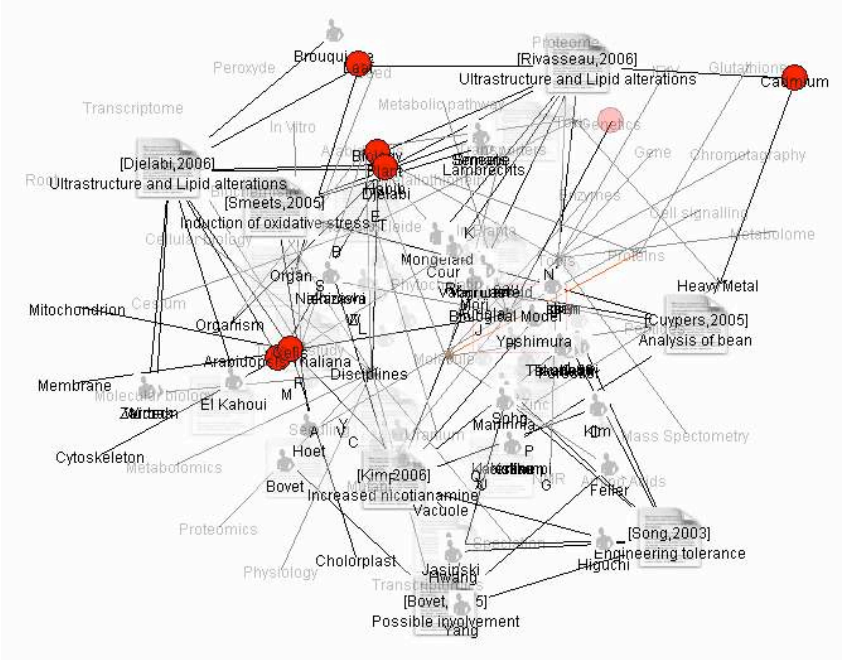

Fig. 1 – Visualisation du référentiel ToxNuc-E sans organisation sémantique

5.1 Visualisation basée sur la sémantique et la topologie du graphe

Nous cherchons à respecter les critères d'esthétisme classiques : limiter le nombre de croisements d'arêtes et la superposition des nœuds. Par ailleurs, nous cherchons à rendre la sémantique du réseau explicite pour faciliter le repérage des entités et la navigation entre entités. Pour atteindre ces objectifs, nous utilisons notre environnement de dessin de graphes Molage et la méthode de construction DVC (Domain-View-Control) basée sur la sémantique du domaine présentée dans (Crampes *et al*., 2006).

Le domaine est modélisé sous forme d'une ontologie. Par ailleurs nous disposons d'une ontologie Molage dans laquelle figurent les types d'entités manipulées : types d'atomes (objets visualisés), types de forces, types de liaisons, ainsi que les attributs qui leur sont associés (couleur de l'objet, forme, position en *x,y*, mouvements autorisés, etc.). On y retrouve aussi les dispositions spatiales possibles : libre, selon un ordre alphabétique, temporel, etc. Nous effectuons une correspondance du domaine vers l'ontologie Molage en spécifiant les valeurs d'attributs. Ensuite il est possible de construire la structure spatiale du référentiel représenté par les objets Molage soit à la main, soit à l'aide d'un scénario automatique comme il sera vu ci-dessous. Finalement, diverses lentilles sont décrites et appliquées pour représenter le contenu de manière typée et sélective en fonction des interactions.

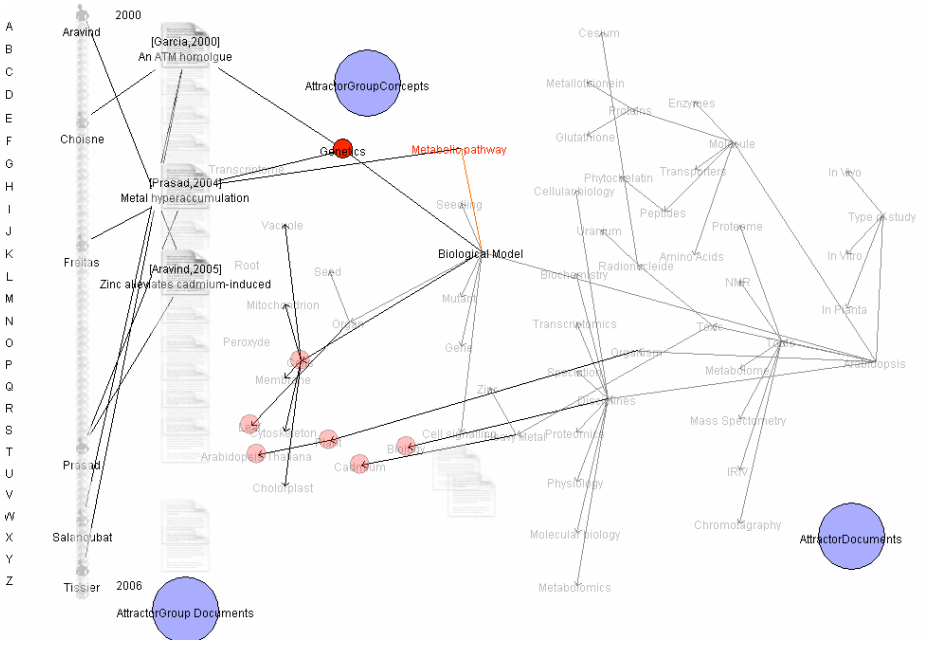

Fig. 2 – Mise en espace du référentiel ToxNuc avec constructeurs (provisoirement) visibles

Une visualisation du référentiel ToxNuc-E réalisée selon cette approche est présentée dans la figure 2. Il est important de noter ici que toute la construction repose sur la mise en œuvre de manière séquencée d'un jeu de trois forces (détaillées dans la section 5.2). Les auteurs sont disposés par ordre alphabétique verticalement à gauche, les documents sont présentés verticalement en vis-à-vis des auteurs pour limiter la surcharge visuelle du graphe bipartite formé par la relation 'auteur-de'. Les concepts sont organisés en arbre de la droite vers la gauche à partir de la racine fixée à droite. On observe la présence temporaire de trois atomes '*constructeurs'* (grosses pastilles) qui servent, à construire le paysage selon sa logique sémantique.

5.2 Réalisation de la visualisation

La disposition de la figure 2 a été réalisée simplement à l'aide d'une succession d'interactions qui déclenchent ou inhibent certaines forces. Pour effectuer des constructions élémentaires, nous utilisons une première force de propriété qui consiste à calculer la distance euclidienne entre deux atomes munis d'un ensemble d'attributs valués. Un ressort est associé aux deux atomes avec pour longueur cette distance. Dans le cadre du projet ToxNuc-E, l'indexation sémantique a fait l'objet d'un travail de recherche particulier ayant débouché sur la méthode d'indexation décrite dans (Shetty *et al.*, 2006). Les entités visualisées ici ont été indexées grâce à cette méthode et leur ensemble d'attributs représente leur score d'intérêt pour chacun des sous-projets de ToxNuc-E. Au déclenchement de l'algorithme de placement FDP, les atomes vont se disposer au plus près de leur distance euclidienne. L'originalité et la

puissance de construction de notre environnement repose sur l'aptitude à ne déclencher cette force que sur un sous-ensemble de propriétés (projection sélective), entre types d'objets présélectionnés, et éventuellement avec des contraintes de déplacement. Nous utilisons aussi une deuxième force de répulsion qui simule la force de Vander Waals en physique ; elle consiste à séparer les atomes à une certaine distance prédéfinie (*limit force*). Une troisième force utilisée est la projection FDP classique pour laquelle la longueur des ressorts est prédéfinie et non calculée à partir d'une sélection de propriétés. Par des combinaisons astucieuses de ces trois forces on peut organiser différemment la base de connaissance.

5.3 Variations dynamiques

On a vu qu'une disposition complexe peut être réalisée simplement par quelques interactions, là ou des algorithmes déterministes et spécifiques à chaque disposition auraient dû être implémentés. Mais le plus grand intérêt de ce jeu de construction interactif réside dans le fait qu'il est possible de passer d'une disposition à une autre par une nouvelle série d'interactions simples. La figure 3 montre une autre vue des mêmes informations, qui résulte d'une suite d'interactions visant à mettre en avant le cœur humain du projet ToxNuc-E et les proximités scientifiques entre les acteurs.

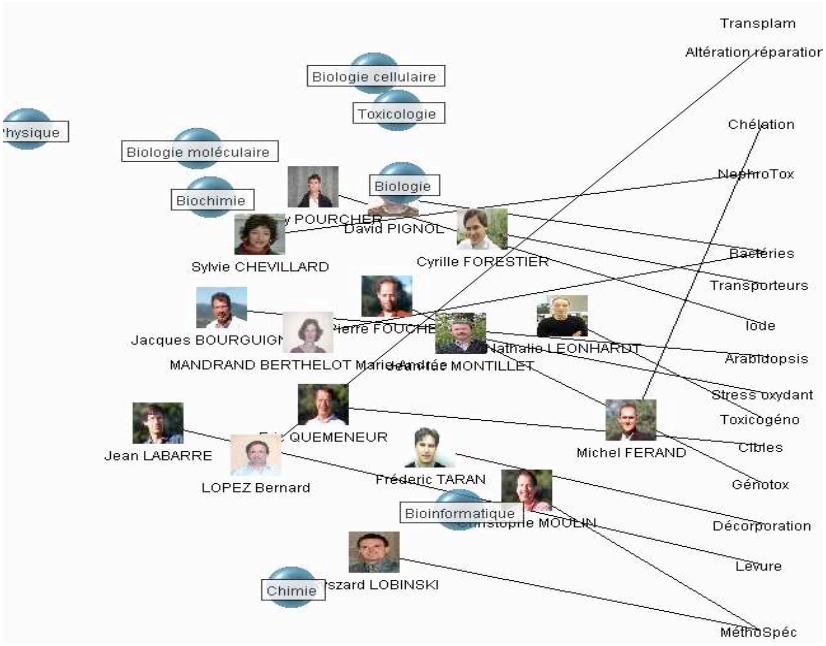

Fig. 3 – La vue humaine du référentiel (extrait)

5.4 Lentilles navigation et navigation basées sur la sémantique

Au final, étant donné le grand nombre d'objets, on ne fait apparaître que l'information utile à un moment précis selon la sémantique (hiérarchie et topologie).

Ceci est réalisé à l'aide de différentes lentilles paramétrables. Par exemple la figure 2 met en avant une 'lentille topologique'. Quand on se place au-dessus d'un concept, les documents indexés par ce concept apparaissent avec la référence, ainsi que les auteurs, les autres documents de ces auteurs, et les autres concepts liés à ces documents. La lentille peut être déformante (*fish eye*) ou transformante (*magic lens*) pour faire apparaître des détails (par exemple les hyperliens vers les documents). Un mode original de navigation sélective est aussi grandement facilité par l'utilisation des forces sémantiques. La liste des propriétés associées aux objets est présentée à gauche de la carte du référentiel. Quand l'utilisateur sélectionne une ou plusieurs propriétés et actionne la force de propriétés, les objets dotés des propriétés sélectionnés se déplacent pour retrouver leur distance sémantique. Ce principe peut être appliqué de multiples façons pour rendre la carte dynamique. Dans la figure 2, trois des propriétés ont été sélectionnées et on voit deux documents qui sont extraits du lot des documents. Ce mode de filtrage selon lequel la carte se transforme en fonction de la sémantique de l'interaction est particulièrement apprécié par les utilisateurs, et nous en approfondissons les multiples variantes qu'il autorise.

6 Conclusion et perspectives

Le concept de référentiel de connaissances paraît très prometteur pour beaucoup d'organismes et de projets, mais il semble que leur construction et leur mise en œuvre butte sur des difficultés. Parmi celles-ci nous voyons le besoin de visualiser un univers complexe et de naviguer facilement et utilement à l'intérieur de cet univers. Pour lever ce verrou, nous avons proposé une approche cartographique dynamique basée sur l'application d'un jeu de forces contrôlées associées à des lentilles. Les résultats sont tout à fait intéressants. L'environnement Molage initialement construit pour gérer des collections musicales (Crampes *et al*., 2006) s'est avéré totalement pertinent pour construire les vues d'un référentiel réel, le projet ToxNuc-E. Les cartes construites ont été présentées à de nombreux acteurs du projet et les perspectives d'utilisation sont réelles. Toutefois, l'intérêt de disposer de plusieurs vues est essentiel, et il faut pouvoir à ce moment là enchaîner différents jeux d'interaction.

Cependant, si la mise en œuvre des interactions ci-dessus est facile quand on connaît les méthodes à appliquer, leur apprentissage est long. De plus le passage d'une vue comme dans la figure 2 à une vue comme dans la figure 3 demande trop d'interactions pour être réellement applicable. Pour répondre au problème, nous considérons une suite d'interactions comme un scénario, c'est-à-dire un enchaînement d'actions conditionnées, une action correspondant à une interaction élémentaire. La construction d'une vue consiste alors à lancer le scénario correspondant, et le passage d'une vue à une autre à lancer un autre scénario. Les scénarios sont écrits en RDF et un moteur les interprète pour les appliquer. Ce langage de scénario et le moteur sont opérationnels dans un état expérimental.

Des scénarios peuvent être lancés en parallèle, mais il existe des risques de conflits. Dans la suite des recherches, un scénario sera une combinaison d'autres scénarios. Un scénario sera aussi représenté par un atome, et une succession de scéna-

rios par une *playlist*. Nous voyons ainsi apparaître un domaine de connaissance particulier qui porte sur la construction dynamique de structures à partir de forces et de scénarios d'application de ces forces.

Références

ARNHEIM R. (1969) *Visual Thinking. Berkeley (California, USA), UC Press*.

ASHBURNER M. & AL. (2000) Gene Ontology: tool for the unification of biology. *Nature Genetics* Vol.25, p. 25 – 29.

BATTISTA G, EADES P, TAMASSIA R, TOLLIS I. (1999). *Graph drawing. Algorithms for the visualisation of graphs. Prentice Hall.*

CARD, S. K., MACKINLAY, J.D. SHNEIDERMAN B. (1999). *Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann publishers, San Francisco, CA, USA.*

CRAMPES M. & RANWEZ S. (2000). Ontology-supported and ontology-driven conceptual navigation on the World Wide Web. *Hypertext 2000.* p. 191-199. San Antonio, Texas.

CRAMPES, M., VILLERD, J., VELICKOVSKI, F., MILLE, N. & RANWEZ, S. (2006) Cartes conceptuelles pour l'ingénierie de cartes de connaissances adaptatives. *Conférence Ingénierie des Connaissances, Nantes*.

EADES P. (1984) A heuristic for graph drawing**.** *In proc. of the 13th Manitoba Conference on Numerical Mathematics and Computing. Vol. 24. Utilitas Mathematica, University of Manitoba, Winnipeg, Canada, Sept-Oct*.

FALOUTSOS C., LIN K., (1995). FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. *Proceedings of ACM SIGMOD, SIGMOD RECORD, vol.24(2), p. 163-174.*

JALABERT F. (2007) *Cartographie des connaissances : l'intégration et la visualisation au service de la biologie Application à l'ingénierie des connaissances et à l'analyse de données d'expression de gènes*. Mémoire de doctorat en informatique de l'université de Montpellier II.

KOHONEN, T. (2001) *Self Organizing Maps. Springer-Verlag, N.Y.*

MACEACHREAN A.M. (1995). *How maps work: Representation, visualization, and design. The Guilford press*.

MARTIN P. A., EBOUEYA Dr M., JO J.H., UDEN L. (2006). Between too informal and too formal. *KMO 2006, International Conference on Knowledge Management in Organizations. Slovenia.*

MARTIN P., BLUMENSTEIN M. & DEER P. (2005).Toward cooperatively-built knowledge repositories. *ICCS 2005, 13th International Conference on Conceptual Structures.* Springer Verlag, LNAI 3596, p. 411-424, Kassel, Germany.

SHETTY R., RICCIO, P. M., QUINQUETON J. (2006). *Hybrid Model for Knowledge Representation. ICHIT 2006, International Conference on Hybrid Information Technology. IEEE.*

SHNEIDERMAN B. (1987) *Designing the User Interface: Strategies for Effective Human-Computer Interaction. Addison-Wesley.*

SHNEIDERMAN (1996). The eyes have it: A task by data type taxonomy for information visualization. *IEEE Workshop on Visual Languages*. Maryland, U.S.A., p. 336-343

ST-AMANT G. & RENARD L. (2004) Présentation d'un premier référentiel des connaissances des capacités organisationnelles de l'administration électronique. *Systèmes d'information : perspectives critiques ; 9ème colloque de l'AIM,* Evry.

TUFTE, E. (1990). *Envisioning Information.* Graphics Press.