



HAL
open science

Recherche exploratoire : proposition d'une méthode basée sur une ontologie de domaine

Gérald Kembellec

► **To cite this version:**

Gérald Kembellec. Recherche exploratoire : proposition d'une méthode basée sur une ontologie de domaine. Contextes, langues et cultures dans l'organisation des connaissances, ISKO France, Oct 2013, Paris, France. pp.281-302. hal-01348474

HAL Id: hal-01348474

<https://hal.science/hal-01348474>

Submitted on 3 Feb 2018

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.

« Recherche exploratoire : proposition d'une méthode basée sur une ontologie de domaine »

Gérald KEMBELLEC
Laboratoire Paragraphe
gerald.kembellec@cnam.fr

Thèmes :

- Contextualisation et visualisation de l'information
- Conception centrée utilisateur et utilisation

Mots clés : sémantique ; recherche exploratoire ; cognition ; système de recherche d'informations ; ontologie

Résumé

Les exigences du public faceaux systèmes de recherche documentaires ne cessent décroître en terme d'assistance, de précision et d'usabilité, alors que dans un même temps la surabondance informationnelle noie l'utilisateur dans un flot de données principalement composé de bruit au travers des moteurs de recherche traditionnels. Comment, dans ce contexte, aider les usagers à résoudre des besoins informationnels spécifiques dus à une connaissance disparate du domaine de connaissance et des méthodes de recherche ? Nous avons souhaité dans cet article dépasser le traditionnel modèle question/réponses des principaux moteurs de recherche avec la recherche exploratoire (*exploratory search*). Cette méthode offre le double avantage de la facilité d'utilisation avec une recherche par spécification et des bénéfices pour l'utilisateur en terme de compréhension et d'apprentissage. Après un rappel historique et un état de l'art, nous introduirons une heuristique de recherche d'information pluri-tactiques que nous avons implémentée dans une interface graphique basée sur l'exploration de la représentation graphique d'une ontologie de domaine : OntologyNavigator.

Introduction

Les exigences relativesaux systèmes de recherche ne cessent décroître, il est plus que jamais nécessaire d'assister l'accès à la connaissance au-delà du simple paradigme de question-réponses. Les processus de recherche informationnelles'inscrivent dans des démarches à la fois opportunistes (M. Bates, 1989), itératives, (Guthrie, 1988) et multitactiques (Marchionini, 1997; Zhang & Marchionini, 2004). Les usagers dans ce contexte visent à résoudre des problèmes complexes, à développer et renforcer leurs capacités cognitives. Malgré la tendance « *informavore*¹ » de notre espèce (Miller, 1983), nous sommes contraints par le volume documentaire de sélectionner l'information pour réduire la surcharge informationnelle. Les méthodes de prétraitement de l'information, comme les systèmes de recommandation des outils de *Social Tagging & Bookmarking* personnalisent les flux informationnels pour en extraire les éléments redondants et inadéquats dans le contexte (Loeb & Terry, 1992). Cependant, l'invasion des tablettes, *smartphones* et autres micro-ordinateurs connectés en continu aux réseaux sociaux et autres flux informationnels en tous genres rend notre attention difficile à mobiliser par ce type d'alertes. En effet, réaliser un système de veille informationnel par alertes (mail, RSS, SMS ou autre) capable de nous enrichir devient un réel

¹En français dans le texte

défi. L'information pertinente issue de la veille est noyée dans un flux constant de bruit informationnel, notre attention étant surexploitée par des alertes communicationnelles sonores et/ou visuelles hautement disruptives pour le processus de recherche avec un bénéfice souvent faible.

Un objectif est né de ces réflexions : dépasser le traditionnel modèle question/réponses des principaux moteurs de recherche. Dès les années 1960, les systèmes de recherche ont été imaginés pour consolider le raisonnement et accroître connaissance et intelligence (Engelbart, 1962; Licklider, 1960). À cette époque, il n'existait pas de technologie supportant la mise en œuvre de tels dispositifs cognitifs exploratoires pour de gros corpus (White & Roth, 2009). Cet article présentera l'évolution historique de la recherche d'information exploratoire, comment ce processus peut faire sens dans la construction du savoir et quelle est la limite de cette méthode. Nous présenterons ensuite quelques modèles de systèmes efficaces de recherche exploratoire, puis nous proposerons *Ontology Navigator*, notre modèle appuyé sur une heuristique de recherche originale.

I. La recherche exploratoire ou « *exploratory search* »

I.1. Aspects historiques

La recherche d'information est couramment perçue comme la localisation et le traitement d'une ou plusieurs informations au sein d'un environnement documentaire complexe, dans le but de répondre à une question ou de résoudre un problème (Dinet & Rouet, 2002). Cette acception du terme part du principe qu'un usager en contexte de recherche d'information a pleine conscience de l'étendue du sujet de sa prospection. Les différents schémas historiques vont en ce sens : le besoin d'information va être comblé grâce à la recherche, le tri, la lecture et la compréhension de documents vecteurs de connaissance (Guthrie, 1988; Kuhlthau, 2005; Marchionini, 1997). Le fonctionnement de ce paradigme était étroitement lié à l'idée qu'il suffisait de formuler une requête au système d'information pour en extraire les documents en adéquation avec le besoin (M. Bates, 1989). Ainsi, il n'était alors pas envisagé que l'utilisateur d'un système d'information n'ait qu'une conscience vague de ce qu'il recherche. Dès 1997, Marchionini avait proposé un schéma du processus de recherche d'informations découpé en étapes distinctes comme la prise de conscience d'une lacune puis la définition-compréhension de la problématique associée (Marchionini, 1997). Dans ce cadre, comment un système pourrait-il fournir des documents à un usager incapable de définir et de formuler son questionnement ? Cette problématique a été envisagée par Kuhlthau qui donne une dimension psychologique au processus informationnel (Kuhlthau, 2004):

« Le processus de recherche d'information est enclenché par un état d'incertitude dû à un manque de compréhension, à un sens inexpliqué, à une structure incomplète. Il s'agit d'un état de nature cognitive qui provoque généralement des symptômes affectifs comme l'anxiété et le manque de confiance. Il faut s'attendre à la présence d'incertitude et d'anxiété dans les premières étapes du processus de recherche d'information. Les symptômes affectifs d'incertitude, de confusion et de frustration sont accompagnés de pensées vagues et floues à propos d'un sujet ou d'une question. »

La disparition des sensations d'angoisse va de pair avec la résolution du manque de connaissance et la compréhension du sujet, ce qui permettra de rechercher sereinement des documents sur un sujet conceptuellement appréhendé.

Pour ces raisons, une nouvelle discipline nommée *Human Computer Information Retrieval*, entre SIC et informatique, est apparue. Elle encourage l'appropriation du système par l'utilisateur au moyen de techniques éprouvées comme le paramétrage de la visualisation et l'appréciation de la pertinence des résultats (Card, Mackinlay, & Shneiderman, 1999; Salton & Buckley, 1997). L'observation des données utilisateur sera utilisée pour lui proposer une assistance à la saisie personnalisée ou des recommandations (Koenemann & Belkin, 1996). Une méthode alternative d'accès aux informations émerge donc sous forme tutélaire, tout en agissant sur les fonctions cognitives de l'intelligence humaine. Ce phénomène, baptisé « recherche exploratoire » est une nouvelle frontière dans le domaine de l'accès à l'information.

I.2. Recherche exploratoire et sémantique

Le mode d'accès à l'information à travers un système onomasiologique dans une taxonomie, n'est pas précisément novateur. Cette méthode a fait, dès la fin du siècle dernier, les grandes heures de sociétés comme Altavista et Yahoo ou encore le projet de portail classificatoire DMOZ. Cette classification généraliste est basée sur le travail documentaire d'évaluateurs volontaires qui passent en revue des sites Web et les classent par thématique. DMOZ a été utilisé pendant des années par Google dans l'optique de faire un « annuaire du web » avant d'être délaissé par le géant de MountainView en 2011.

Il peut de prime abord sembler étrange, voire rétrograde, de revenir sur un modèle d'accès à l'information qui est en perte de vitesse depuis une décennie. Cependant, il faut mettre en perspective le fait qu'à l'époque de l'âge d'or des annuaires du web la classification des sites web était un travail éditorial manuel consciencieux. Ce qui a fait la force de ces annuaires, la classification humaine, est devenu une faiblesse à cause de l'inévitable silence provoqué par la surabondance informationnelle que ne pourrait rattraper une armée de documentalistes en *crowdsourcing*.

Cependant avec le tournant du siècle, l'idée de sémantiser les pages web s'est imposée au point de devenir un facteur d'indexation pour les nouveaux moteurs des grands moteurs. Avec Panda, nouvel algorithme d'indexation de l'Internet, Google a imposé à l'été 2011 un changement de paradigme dans la recherche d'information. Avec sa logique orientée web sémantique, la méthode Panda a révolutionné l'accès aux données numériques. En effet, la valeur des sites web n'est plus attribuée uniquement grâce aux « *backlinks* » (liens entrants) et à l'analyse statistique du contenu, mais aussi sur la cohérence et la richesse des règles de description. L'algorithme a donc été modifié pour prendre davantage en compte le contenu et la sémantique des sites web pour le classement d'une page par rapport à une thématique donnée. Les documents dont les métadonnées s'accordent avec les formats du Web de données sont largement privilégiés. Dans ce contexte, les méta-informations sont encapsulées au sein des documents hypertextes selon des protocoles dits « web 3.0 » ou web de données, comme le RDFa et les micro-formats. Ainsi, la structuration et l'annotation des documents offrent aux outils d'indexation et d'interrogation de sélectionner, filtrer une information plus en adéquation avec la requête de l'utilisateur (Benali, Rieu, & Soule-Dupuy, 2009). Les données et métadonnées des pages web devenant liées à des vocabulaires contrôlés au travers des grammaires formalisées, formant des réseaux sémantiques, sont contrôlées par des ontologies qui permettent théoriquement de classer le corpus du web dans un graphe conceptuel complexe.

I.3. État de l'art de la recherche exploratoire

La recherche d'informations implique des contextes à la fois humains et technologiques, conciliant dans une même interface le besoin de simplicité induit par les capacités humaines et la complexité requise par l'accès à l'information (Belkin, Oddy, & Brooks, 1982). Certains systèmes de recherche vont dans cette optique jusqu'à offrir des interfaces dépourvues de formulaire de requête textuelle classique. Cette méthode présente le double avantage de ne pas autoriser de saisie « hors champ » et de rester ainsi dans le corpus indexé évitant ainsi de mettre l'utilisateur en situation d'échec face au silence.

Certains environnements de recherche ont exploité le modèle Query By Example² (QBE) de Zloof qui exploite la catégorisation et le thésaurus pour proposer des résultats de plus en plus fins dans un corpus (Zloof, 1977). Ce modèle, particulièrement adapté à la vente en ligne a été implémenté dans de nombreux sites commerciaux comme Amazon pour faciliter la sélection puis l'achat d'un produit.

Chalmers et Chitson ont proposé dès 1992 une interface graphique en trois dimensions pour la visualisation/navigation de corpus scientifiques au travers de plusieurs critères (Chalmers & Chitson, 1992). Cependant, cette méthode n'est que partiellement efficace, car elle n'offre pas la possibilité de filtrage fin et rapide : le bruit reste important et à moins d'utiliser un système classificatoire à forte granularité elle est d'un usage chronophage. Avec le Visual Catalog, Papy, Chauvin et Leblond alignaient la classification Dewey et le thésaurus Rameau pour créer un mode de navigation exploratoire thématique du catalogue d'une bibliothèque universitaire (Papy & Chauvin, 2005; Papy & Leblond, 2007). Le prototype expérimental ASE montre de manière efficiente l'intérêt de combiner, dans une même interface thématique de recherche documentaire, statistiques et contextualisation des citations pour le filtrage et la sélection interactive des documents clés (Dunne, Shneiderman, Gove, Klavans, & Dorr, 2012). Cependant, de l'aveu même des auteurs, ASE exige actuellement un « substantiel traitement de données » en amont et la plupart des corpus nécessaires ne sont pas disponibles à partir des bases de connaissances des éditeurs. Plus récemment, le système Adaptive VIBE (Ahn & Brusilovsky, 2013) a proposé une interface de cartographie exploratoire d'une thématique de recherche au moyen de mise en exergue, de catégorisation et de regroupement de « points d'intérêts » paramétrables. Dans la même veine, le projet HdaLab³ utilise propose une interface graphique basée sur la catégorisation et les ressources sémantisées de Wikipédia pour présenter l'histoire de l'art à travers un portail dont les ressources sont accessibles par recherche exploratoire au travers d'une taxonomie mais aussi par facettes ou thésaurus (Sajus, s. d.; Scoffoni et al., 2012).

I.4. Visualisation, interaction et recherche exploratoire

Un portail documentaire tourné vers la recherche exploratoire doit se doter d'une interface avec laquelle l'utilisateur doit interagir pour explorer et appréhender un corpus.

Tricot et Roche reprenant Card, Mackinlay et Shneiderman répartissent la présentation navigable de l'information en trois types de paradigmes cartographiques (Card et al., 1999; Tricot & Roche, 2006; Tricot, 2006):

– Les paradigmes de représentation qui définissent la structure informationnelle :

1. la structure tabulaire (Rao & Card, 1994) ;
2. la structure arborescente (Van Ham & van Wijk, 2003) ;
3. la structure par graphe (Cassidy, Walsh, & Coghlan, 2006) ;

² Requête par l'exemple

³<http://hdaLab.iri-research.org/hdaLab/categories/>

4. la structure en spirale (Carlis & Konstan, 1998)

– Les paradigmes de visualisation représentent les informations d'une manière claire et cohérente dans un espace physique limité :

1. Les techniques de visualisation uniformes avec une vue d'ensemble et les détails à la demande (Jerding & Stasko, 1998).
2. Les techniques de visualisation hétérogènes en focus avec contexte (Card et al., 1999).

– Les paradigmes d'interaction permettent aux utilisateurs d'interagir avec la visualisation :

1. la navigation multi échelle – zoom / panoramique (Perlin & Fox, 1993);
2. le focus et le contexte – listes hiérarchiques zoomables (Card et al., 1999);
3. le filtrage dynamique (Pediatakis & Hascoët-Zizi, 1996) ;
4. le zoom sémantique (Bartram, Ho, Dill, & Henigman, 1995; Hascoët & Beaudouin-Lafon, 2001).

Modéliser un système documentaire permettant de décrire le sujet et le contexte de recherche associé sans perdre de vue le contexte global d'un domaine de connaissance tout cela dans une même interface représente un défi ergonomique. La seconde partie de cet article va présenter un modèle graphique de recherche d'information alliant les recherches exploratoire et traditionnelle pour optimiser l'expérience navigation et son bénéfice pour l'utilisateur.

II. Notre modèle : **Ontology Navigator**

Enrichi des expériences et théories des chercheurs en information-documentation et psychocogniticiens, nous avons tenté de modéliser et de réaliser un système de recherche exploratoire qui guiderait l'utilisateur dans sa recherche tout en lui proposant des informations annexes sur le sujet.

Postulant que les usagers possédant des habiletés avec les technologies et techniques de recherche d'information ont autant de chance de tirer de l'information pertinente d'un système d'information que les spécialistes du domaine de connaissance (Ihadjadene & Martins, 2004), comment penser une interface capable de répondre aux besoins de ces deux catégories ? Par ailleurs, les utilisateurs qui ne sont à l'aise ni avec la méthodologie de recherche d'informations, ni avec le domaine, n'ont pas la compétence pour formuler des requêtes ou naviguer dans des espaces informationnels complexes.

II.1. Proposition d'un modèle

Les systèmes de recherche exploratoire répondent à ces attentes en s'appuyant sur des relations homme-machine symbiotiques qui fournissent une assistance à l'exploration de paysages informationnels inconnus. Selon Zhang, il y a trois paradigmes de visualisation en contexte de recherche d'informations (Zhang, 2008):

1. *Query searching and Browsing*. Une requête limite les résultats présentés sous forme visuelle pour que les utilisateurs la parcourent et concentrent leur champ visuel sur des informations spécifiques.
2. *Browsing and Query searching*. Une présentation visuelle de l'ensemble du domaine informationnel est proposée, les utilisateurs parcourent le système et les résultats sont mis en évidence.
3. *Browsing Only*. Il s'agit uniquement de navigation, sans composante de requête.

Pour résoudre le problème posé par l'accès direct à l'information, il faut rejoindre le paradigme *Querysearchingand Browsing* de Zhang. Cette solution n'est que partiellement pertinente puisqu'elle annule les bienfaits cognitifs de la navigation. Nous avons donc proposé d'étendre le modèle de Zhang avec une heuristique de recherche supplémentaire *Query, Browsing, Query-Searching* (Kembellec, 2012) :

1. Une requête par auto-complétion permet de découvrir un contexte informationnel dans une taxonomie du domaine de connaissance.
2. Un focus se crée autour ce contexte permettant d'affiner par navigation et lancer la fonction de recherche de documents associés. Si le terme ou l'entité nommée que l'utilisateur cherchait n'est pas répertorié, libre à lui de l'ajouter.
3. Enfin, les concepts trouvés qui sont en adéquation avec le besoin informationnel vont être utilisés pour interroger les bases documentaires spécialisées du domaine de connaissance en combinant terminologie et indice classificatoire.

L'utilisateur aura découvert lors de son usage du système documentaire des sujets connexes par sérendipité et aura personnalisé le point d'accès pour son bénéfice et enrichi par sa participation la base commune de termes descripteurs pour l'enrichissement commun selon le principe des folksonomies.

Notre modèle théorique *Query, Browsing, Query-Searching* (Kembellec, 2012) se décompose en étapes simples itératives et pluri-tactiques comme proposé par Guthrie, Marchionini et Zhang (Guthrie, 1988, Marchionini, 1997; Zhang & Marchionini, 2004) en espérant que les étapes provoquent un déclic opportuniste de sérendipité chez l'utilisateur qui l'enrichira cognitivement et informationnellement au-delà de sa demande initiale (M. J. Bates, 1990; M. Bates, 1989).

II.2. Implémentation du modèle

Nous avons décidé de mettre à l'épreuve notre heuristique « *Query, Browsing, Query-Searching* » en créant une interface capable de respecter l'accès pluri-tactique à la documentation. Le cadre de ce modèle est l'information scientifique et technique en informatique. L'avantage de la modélisation de ce domaine scientifique en particulier est qu'il existe déjà un système taxonomique et un thésaurus mondialement reconnu par les scientifiques, ingénieurs et sociétés savantes : la Computer Classification System ou CCS éditée par l'Association for Computing Machinery (ACM). Cette société savante maintient et fait évoluer la classification et les vocabulaires contrôlés associés depuis 1964⁴. La CCS est utilisée pour indexer le corpus documentaire de plus de deux millions d'articles scientifique proposé par l'ACM⁵, sa dernière révision majeure date d'octobre 2012.

Dans l'état de l'art, nous avons évoqué l'alignement de la classification Dewey avec le thésaurus Rameau dans l'optique d'une navigation exploratoire thématique du catalogue d'une bibliothèque universitaire (Papy & Chauvin, 2005; Papy & Leblond, 2007). Nous avons pris le même principe avec un changement d'échelle. Après avoir recensé les principales bases de connaissances scientifiques mondiales du domaine informatique, nous avons réalisé un alignement des vocabulaires contrôlés ACM (entités nommées et termes) avec la taxonomie éponyme et les langages balisés de description de données. Nous avons ainsi créé une ontologie du domaine informatique qui inventorie « virtuellement » les contenus des grandes bases de données scientifiques et techniques en créant des requêtes spécifiques

⁴<http://www.acm.org/about/class/2012>, accédé le 1 juillet 2013

⁵<http://dl.acm.org>, accédé le 1 juillet 2013

dynamiquement hyperliées au concept mis en focus (cf. Figure 2) en utilisant selon le cas un indice classificatoire ou des éléments de vocabulaire contrôlé.

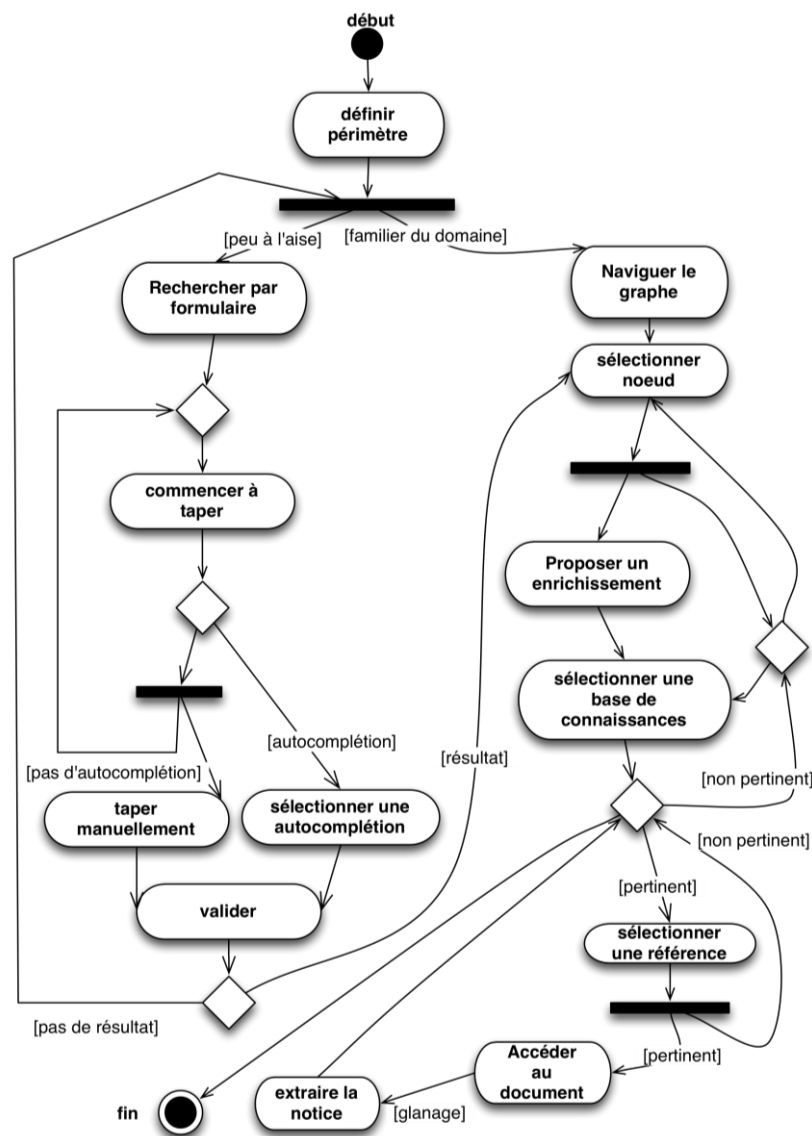


Figure 1. Explication de notre modèle de recherche documentaire

Pour créer l'interface de visualisation cartographique du domaine de connaissance, nous avons choisi le paradigme de représentation par graphe qui fait consensus (Cassidy et al., 2006; Tricot & Roche, 2006; Tricot, 2006). Pour atteindre notre objectif cognitif, notre choix de visualisation s'est porté sur la méthode hétérogène « *focus + context* » qui offre à la fois le détail du choix et le contexte général (Card et al., 1999) ce qui nous a amené à adopter le même choix pour l'interaction. Cependant, nous avons choisi de doubler le processus de sélection/interaction d'un formulaire de recherche par auto complétion dont l'usage peut s'apparenter à un filtrage dynamique balisé du thésaurus. En effet, lettre après lettre, l'utilisateur peut affiner sa recherche balisée par les termes et entités nommées proposées.

Afin de ménager la capacité psycho-cognitive des usagers, nous ne présumerons pas des aptitudes documentaires et du niveau de familiarité avec le domaine de connaissance de l'utilisateur (Ihadjadene & Martins, 2004). Le diagramme en figure 1 explicite les méthodes

d'accès à de l'information scientifique et technique dans un domaine de connaissance balisé dans une ontologie.

Un utilisateur peu à l'aise avec le sujet sur lequel il souhaite se renseigner va sans doute préférer utiliser le formulaire assisté par auto-complétion. Ce formulaire va utiliser le moteur sémantique pour positionner la requête dans une taxonomie du domaine de connaissance grâce aux termes ou entités nommées et mots clés (respectivement issus des vocabulaires contrôlés et mots clés libres). L'avantage d'utiliser l'auto-complétion dans le formulaire de requête est de ne permettre que des requêtes qui permettront de se positionner correctement dans la taxonomie sans passer par l'étape décourageante d'absence de résultats. Une fois un terme ou une entité nommée sélectionnée et son contexte désambiguïsé une représentation graphique permet de le contextualiser dans le domaine de connaissance et offre également un focus sur le concept lié. L'utilisateur pourra ainsi comprendre la place du concept dans la taxonomie et le vocabulaire associé. Il pourra alors choisir d'aller chercher de la littérature associée au concept sur une des bases de connaissances associées et ce par un simple clic sur un hyperlien, la requête étant générée par le système à partir du contexte en collaboration avec les « *webservices* » des bases documentaires tels que l'OAI-PMH, les *endpoints* SPARQL ou encore le Z39-50 (Kembellec, 2012, p. 223-259 chapitre 9) de rebondir par sérendipité sur un autre concept proche (généralisation/spécification) ou associé (équivalence).

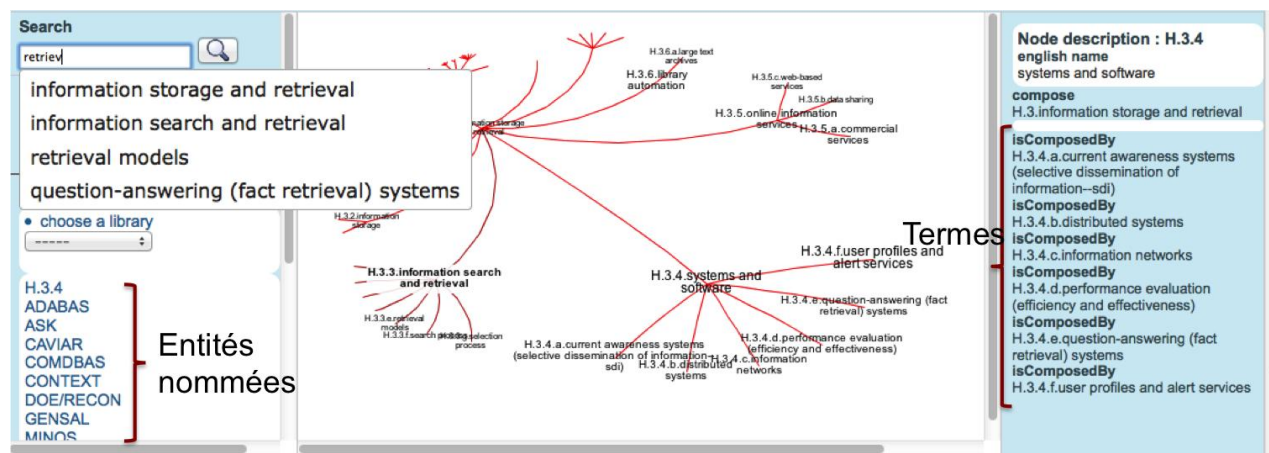


Figure 2. L'interface ontologyNavigator du Modèle QB-QS

Un usager plus à l'aise avec le domaine de connaissance pourra utiliser la deuxième méthode de positionnement taxonomique par spécification dans le graphe de connaissance. Cette démarche est plus enrichissante, car à chaque étape de spécification conceptuelle, l'utilisateur a un aperçu du contexte sémantique. La répétition de l'affichage des termes et entités nommées liés à une branche du graphe de connaissance entrainera automatiquement un bienfait cognitif. De plus, l'usager peut choisir d'enrichir la base de termes et d'entités nommées relativement à un concept s'il constate une lacune. Cette ajout lui permet non seulement de s'appropriier l'outil pour une recherche ultérieure, mais offre un point d'entrée supplémentaire pour un utilisateur novice qui fait une recherche par auto-complétion.

II.2.Evaluation des bienfaits

Historiquement, il a été d'usage de juger de la pertinence algorithmique d'un système de recherche d'informations plutôt que de la qualité de l'interaction entre l'interface et l'humain, plus précisément de la symbiose et des bienfaits qui en découlent (White & Roth, 2009, p. 61-69). L'évaluation des bénéfices de l'interaction homme-machine sur les résultats se fait principalement sur les retours tangibles, à savoir sur le bruit et le rappel présentés par les résultats retournés relativement à la collection (Allan, 2003). Or selon White et Toth, dans le

cas de la recherche exploratoire, il est également nécessaire de quantifier les bénéfices en terme de cognition. Pour mémoire, des psycho-cogniticiens ont démontré très tôt la supériorité rechercheexploratoire sur la recherche traditionnelle(Russell, Stefik, Pirolli, & Card, 1993).

Nous avons soumis notre modèle à la grille d'analyse proposée par White & Roth puis l'avons confronté à un panel d'élèves ingénieurs dans le domaine de l'informatique (après un semestre de cours en recherche de documentation scientifique et technique) et avons étudié leurs retours, résultats que nous allons expliciter sous plusieurs axes.

La mesure du succès d'une tâche de recherche d'information ne doit pas seulement être conditionnée par le fait qu'utilisateur atteigne ou non un document cible en particulier. Il faut également prendre en compte la quantité d'information détaillée rencontrée lors du processus de recherche. En fait, l'intérêt d'une recherche exploratoire dépend autant de la richesse de l'expérience d'utilisation que des résultats documentaires. La compréhension du sujet par la navigation du graphe d'un domaine de connaissance est une métrique importante du processus d'évaluation.

Le temps passé pour atteindre un état de complétude de la tâche peut être un bon moyen d'évaluer l'efficacité des activités d'exploration si on le met en concurrence avec une méthode de recherche traditionnelle. Cependant, comme notre approche est pluri-tactique, le temps passé est difficile à comparer avec une recherche traditionnelle dont la période de compréhension du sujet n'est pas quantifiable.

Si l'on s'appuie sur la spécialiste de l'interaction homme-machine en matière de documentation en lien avec la psychologie cognitive et la communauté de l'éducation il faut également mesurer les bénéfices de l'expérience de recherche en matière psycho-cognitive sur l'utilisateur(Kuhlthau, 1993, 2005).

Après l'atteinte des résultats les bénéfices en termes d'apprentissage, la richesse, l'exhaustivité de la perspective post-exploration d'un utilisateur en termes d'apprentissage et de cognition n'est malheureusement pas mesurable en dehors du retour subjectif de l'utilisateur sur les états de satisfaction et de complétude de compréhension du sujet. Nous proposerons une étude ultérieure dans l'optique de baliser et valider avec des psycho-cogniticiens un protocole scientifique factuel démontrant le réel bénéfice en terme d'apprentissage et de compréhension par l'utilisation de notre modèle *Query, Browsing, Query-Searching* par rapport au strict formulaire ou à la recherche exploratoire seule.

Dans l'attente de ces résultats, nous proposons de comparer la pertinence des documents retrouvés dans les bases documentaires informatiques par les élèves ingénieurs. Le protocole était de trouver dix documents en rapport avec un sujet libre dont l'élève ingénieur est spécialiste. Nous avons également proposé à notre panel de choisir le modèle d'accès à l'information qui leur paraissait le plus adapté à leur besoin d'information avant et après avoir testé l'outil ontologyNavigator. Les différentes interfaces présentées étaient les suivantes :

- le formulaire traditionnel de la bibliothèque numérique ACM ;
- le formulaire avancé de la bibliothèque numérique ACM ;
- une représentation tabulaire de la taxonomie ACM ;
- le projet ontologyNavigator.

Tableau 1. Résultats du choix d'une interface de recherche

Choix d'un Modèle de recherche	Choix à priori	Choix a posteriori
Moteur de recherche spécialisé simple	0 %	0 %
Moteur de recherche spécialisé avec champs avancés	25 %	17 %
Représentation tabulaire	50 %	0 %
OntologyNavigator	17 %	58 %
Ne se prononcent pas	25 %	25 %

Tableau 2. Résultats de la pertinence des documents trouvés par ontologyNavigator

Pertinence des documents trouvés de 1 à 10	Ventilation des évaluations
1-2	0 %
3-4	25 %
5-6	0 %
7-8	50 %
9-10	25 %

Ces deux évaluations de l'outil, l'une portant uniquement sur la qualité des résultats (évaluation expert), l'autre sur l'usabilité perçue puis ressentie. Les résultats mettent en exergue les points suivants :

1. La qualité de l'information retrouvée est satisfaisante mais perfectible.
2. L'interface d'ontologyNavigator est appréciée (choix de 58 % du panel après utilisation). Mais, jugée d'un abord complexe (l'interface est choisie *a priori* par seulement 17 % du panel) dans les commentaires libres, elle ne fait donc pas consensus.

Ces appréciations critiques sont encourageantes notamment par le fait que les utilisateurs (familiers du domaine de connaissance) adoptent notre modèle en délaissant le paradigme traditionnel de requête-réponses. Aucun usager n'a souhaité enrichir la base de mot clés, ce qui pose la question de la masse critique d'utilisateurs avant de pouvoir parler de d'enrichissement collaboratif. Les évaluations négatives permettront de corriger sur les prochaines versions d'ontologyNavigator des faiblesses ergonomiques qui nuisent à l'appropriation de l'outil.

III. Discussion

La réelle efficacité d'un modèle de recherche exploratoire dépend de la capacité d'aligner le système classificatoire avec le corpus et ses métadonnées, ce qui n'est souvent pas possible (Ahn & Brusilovsky, 2013). Il faut espérer que les bases de connaissances documentaires scientifiques en ligne aux métadonnées ouvertes comme ArXiv, Hal, DBLP et Isidore vont ouvrir la voie aux éditeurs scientifiques qui verront dans l'ouverture de leurs métadonnées un meilleur accès à leur catalogue. Il reste à trouver un point d'équilibre entre accessibilité et viabilité du modèle économique.

Malgré l'encouragement aux bonnes pratiques de sémantisation, pour l'instant, peu de bases de connaissance sont suffisamment sémantisées pour être exploitables dans un contexte d'open data. Le travail d'« auto-description normalisée » est pour l'instant loin d'être la règle lors de la création d'un document hypertexte sur Internet, même si certains outils de gestion de contenu commencent à proposer des plug-in en ce sens. Cependant, dès que la sémantisation aura atteint sa masse critique d'exploitation, il est plus que probable de revoir des portails documentaires fleurir sur l'internet. En outre, il est raisonnable de penser que de

nouveaux modes de visualisation vont apparaître pour optimiser la recherche exploratoire non seulement sur les corpus textuels, stimulant nos fonctions cognitives tout en élargissant nos champs de recherche par sérendipité, mais aussi sur des bases de média audio visuelles, pour peu que les règles de sémantisation soient respectées. Notre modèle ontologyNavigator pourrait sortir de son domaine de recherche unique balisé sur un corpus fini pour s'aventurer sur de plus vastes étendues de connaissance.

IV. Conclusion

Après avoir présenté la recherche exploratoire comme alternative au modèle traditionnel de recherche d'information au moyen de formulaire classique, nous avons établi grâce à la littérature les indiscutables bénéfices en terme de cognition sur l'utilisateur notamment grâce à la sémantisation du web. L'état de l'art nous a permis de dégager des évolutions et tendances en terme de projets liés à la recherche exploratoire, ce qui couplé à une étude des méthodes de représentation de l'information sous forme de paradigmes cartographiques nous a permis de proposer une heuristique de recherche exploratoire pluri-tactique baptisée « *Query, Browsing, Query-Searching* ». Ce modèle a ensuite été implémenté dans une interface graphique représentant une ontologie de domaine navigable liant sémantiquement les concepts entre eux et avec les différents vocabulaires contrôlés pour créer des requêtes vers les différentes bases de connaissances offrant des « *webservices* » sur le principe de l'« *open data* ». Enfin nous avons posé les limites de ce système, principalement liées à la faible sémantisation des bases documentaires en ligne.

Bibliographie

- Ahn, J., & Brusilovsky, P. (2013). Adaptive visualization for exploratory information retrieval. *Information Processing & Management*. doi:10.1016/j.ipm.2013.01.007
- Allan, J. (2003). High accuracy retrieval from documents (HARD) track overview. In *Proceedings of the Twelfth Text REtrieval Conference (TREC)* (p. 24–37).
- Bartram, L., Ho, A., Dill, J., & Henigman, F. (1995). The continuous zoom: A constrained fisheye technique for viewing and navigating large information spaces. In *Proceedings of the 8th annual ACM symposium on User interface and software technology* (p. 207–215).
- Bates, M. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, (13), 407–424.
- Bates, M. J. (1990). Where should the person stop and the information search interface start? *Information Processing & Management*, 26(5), 575–591.
- Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982a). ASK for information retrieval: Part I. Background and theory. *Journal of documentation*, 38(2), 61–71.
- Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982b). ASK for information retrieval: Part II. Results of a design study. *Journal of documentation*, 38(3), 145–164.
- Benali, K., Rieu, D., & Soule-Dupuy, C. (2009). *Documents annotés et langages d'indexation* (Hermès-Lavoisier., Vol. 12). Paris: Hermès-Lavoisier.
- Card, S. K., Mackinlay, J., & Shneiderman, B. (Éd.). (1999). *Readings in Information Visualization: Using Vision to Think* (1^{re} éd.). Morgan Kaufmann.
- Carlis, J. V., & Konstan, J. A. (1998). Interactive visualization of serial periodic data. In *Proceedings of the 11th annual ACM symposium on User interface software and technology* (p. 29–38). Présenté à ACM symposium on User interface software and technology, ACM.
- Cassidy, K., Walsh, A., & Coghlan, B. (2006). Using Hyperbolic Geometry for Visualization

of Concept Spaces for Adaptive eLearning. In *Proceedings of the 1st International Workshop on Authoring of Adaptive & Adaptable Hypermedia*. Présenté à 1st Inter. Workshop on Authoring of Adaptive & Adaptable Hypermedia, Dublin, Ireland.

Chalmers, M., & Chitson, P. (1992). Bead: Explorations in information visualization. In *Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval* (p. 330–337).

Dinet, J., & Rouet, J. F. (2002). La recherche d'information: processus cognitifs, facteurs de difficultés et dimension de l'expertise. *Interaction homme-machine et recherche d'information*, 133- 161.

Dunne, C., Shneiderman, B., Gove, R., Klavans, J., & Dorr, B. (2012). Rapid understanding of scientific paper collections: Integrating statistics, text analytics, and visualization. *Journal of the American Society for Information Science and Technology*, 63(12), 2351–2369.

Engelbart, D. (1962). *Augmenting human intellect: a conceptual framework* (Rapport militaire No. AD 289565) (p. 64–90). Arlington, Virginia: Stanford University, Department of Computer Science.

Guthrie, J. T. (1988). Locating information in documents: examination of a cognitive model. *Reading Research Quarterly*, 23(2), 178–199.

Hascoët, M., & Beaudouin-Lafon, M. (2001). Visualisation interactive d'information. *Revue I3*, 1(1), 77–108.

Ihadjadene, M., & Martins, D. (2004). Experts dans le domaine, experts en Internet: Les effets sur la recherche d'information. *Hermès (Paris. 1988)*, 2004, 39, fascicule thématique « Critique de la raison numérique ».

Jerding, D. F., & Stasko, J. T. (1998). The information mural: A technique for displaying and navigating large information spaces. *Visualization and Computer Graphics, IEEE Transactions on*, 4(3), 257–271.

Kembellec, G. (2012). *Bibliographies scientifiques: de la recherche d'informations à la production de documents normés* (Doctorat). Université Paris 8, Saint-Denis.

Koenemann, J., & Belkin, N. J. (1996). A case for interaction: a study of interactive information retrieval behavior and effectiveness. In *Proceedings of the SIGCHI conference on Human factors in computing systems: common ground* (p. 205–212).

Kuhlthau, C. C. (1993). A principle of uncertainty for information seeking. *Journal of Documentation*, 49(4), 339–355.

Kuhlthau, C. C. (2004). Seeking meaning: A process approach to library and information services.

Kuhlthau, C. C. (2005). Information search process. *Hong Kong, China*, 7.

Licklider, J. C. R. (1960). Man-computer symbiosis. *Human Factors in Electronics, IRE Transactions on*, (1), 4–11.

Loeb, S., & Terry, D. (1992). Information filtering. *Communications of the ACM*, 35(12), 26–28.

Marchionini, G. (1997). *Information seeking in electronic environments*. Cambridge University Press.

Miller, G. A. (1983). Informavores. In F. Machlup & U. Mansfield (Éd.), *The Study of Information: Interdisciplinary Messages* (Wiley., p. 111–113). New York, NY, USA: Wiley-Interscience.

- Oddy, R. N. (1977). Information retrieval through man-machine dialogue. *Journal of documentation*, 33(1), 1–14.
- Papy, F., & Chauvin, S. (2005). Pour une approche visuelle et ergonomique dans la recherche et l'exploration d'informations au sein d'un OPAC de SCD. L'exemple du Visual'Catalog. Papy, F., éditeur: *Les bibliothèques numériques*. Hermès, Paris.
- Papy, F., & Leblond, C. (2007). L'interface de recherche d'information du Visual...Catalog : un outil innovant à « double détente ». *Documentaliste-Sciences de l'Information*, Vol. 44(4), 288- 298. doi:10.3917/docs.444.0288
- Pediatakis, N., & Hascoët-Zizi, M. (1996). Visual relevance analysis. In *Proceedings of the first ACM international conference on Digital libraries* (p. 54–62). New York, NY, USA: ACM. doi:10.1145/226931.226942
- Perlin, K., & Fox, D. (1993). Pad: an alternative approach to the computer interface. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques* (p. 57–64).
- Rao, R., & Card, S. K. (1994). The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information. In *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence* (p. 318–322).
- Russell, D. M., Stefik, M. J., Pirolli, P., & Card, S. K. (1993). The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* (p. 269–276).
- Sajus, B. (s. d.). Le tagging sémantique du corpus Histoire des arts : à la croisée du 2.0 et du 3.0. *Documentaliste-Sciences de l'Information*, 49(2), 12- 15.
- Salton, G., & Buckley, C. (1997). Improving retrieval performance by relevance feedback. *Readings in information retrieval*, 355–364.
- Scoffoni, P., Ogez, E., Dubost, L., Maisonneuve, M., Asselin, E., Sajus, B., & Maisonneuve, M. (2012). Méthodes, techniques et outils. *Documentaliste-Sciences de l'Information*, Vol. 49(2), 8- 15. doi:10.3917/docs.492.0008
- Tricot, C. (2006). *La cartographie sémantique, des connaissances à la carte* (Doctorat). Université de Savoie.
- Tricot, C., & Roche, C. (2006). Visualisation of Ontology: a focus and context approach. InSciT2006.
- Van Ham, F., & van Wijk, J. J. (2003). Beamtrees: Compact visualization of large hierarchies. *Information Visualization*, 2(1), 31–39.
- White, R. W., & Roth, R. A. (2009). Exploratory Search: Beyond the Query-Response Paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1), 1- 98. doi:10.2200/S00174ED1V01Y200901ICR003
- Zhang, J. (2008). *Visualization for information retrieval* (1^{re} éd., Vol. 23). Springer.
- Zhang, J., & Marchionini, G. (2004). Coupling browse and search in highly interactive user interfaces: a study of the relation browser++. In *Digital Libraries, 2004. Proceedings of the 2004 Joint ACM/IEEE Conference on* (p. 384).
- Zloof, M. M. (1977). Query-by-example: A data base language. *IBM systems Journal*, 16(4), 324–343.