



HAL
open science

Consultation de documents et sémantique: Application à des publications savantes

Aurélien Bénel, Sylvie Calabretto, Andrea Iacovella, Jean-Marie Pinon

► To cite this version:

Aurélien Bénel, Sylvie Calabretto, Andrea Iacovella, Jean-Marie Pinon. Consultation de documents et sémantique: Application à des publications savantes. 2d colloque international francophone sur l'écrit et le document (CIFED 2000), 2000, Lyon, France. pp.271-280. hal-02954839

HAL Id: hal-02954839

<https://utt.hal.science/hal-02954839v1>

Submitted on 18 Nov 2021

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.

CONSULTATION DE DOCUMENTS ET SEMANTIQUE

APPLICATION A DES PUBLICATIONS SAVANTES

Aurélien Bénel^{**}, Sylvie Calabretto^{*},
Andréa Iacovella⁺, Jean-Marie Pinon^{*}

^{*} Laboratoire d'Ingénierie des Systèmes d'Information. INSA, Lyon, France.
Prenom.Nom@lisi.insa-lyon.fr

⁺ Ecole Française d'Athènes. Athènes, Grèce.
Prenom.Nom@efa.gr

RESUME

Parce qu'il apparaît que le pari en recherche d'information consiste à réduire l'écart entre le niveau du *contenu* des informations et celui de leur *forme*, nous nous intéressons ici à une théorie linguistique de la sémantique et choisissons de placer notre travail dans ce cadre. Au fil de cet article, nous sommes amenés à identifier la recherche de documents comme une navigation entre corpus et à définir ce qui relève de l'interactivité et ce qui relève du traitement informatique. Nous intégrons la recherche de documents dans une activité plus large, plus complexe, impliquant d'avantage l'utilisateur et comprenant entre autres l'enrichissement de la description des documents (simples ou structurés) et le partage de cette description avec d'autres utilisateurs. Le but consiste en effet à assister le lecteur de publications savantes en le soulageant des aspects répétitifs de son activité et en le laissant se concentrer sur les aspects créatifs, intuitifs et à haut niveau d'abstraction.

MOTS-CLEFS

Recherche d'information, Indexation, Réseaux sémantiques, Lecture/Annotation, Collaboration, Navigation.

1. PROBLEMATIQUE

Par recherche d'information qu'entendons-nous ? Nous reprendrons pour définition celle, aujourd'hui largement admise, de C. J. Van Rijsbergen dans son article [1] de 1986. Par cet article, il souhaitait renouveler le domaine dont les progrès lui semblait insatisfaisants et les limites théoriques atteintes. Sans citer l'intégralité de la définition, nous tenons à faire remarquer qu'elle repose à la fois sur des *correspondances* et des *écarts* (fig.1).

En effet, selon cette définition, la recherche d'information consiste à mettre en correspondance le besoin d'information d'un utilisateur avec des documents (appartenant à une collection) jugés pertinents. Cette mise en correspondance, afin d'être automatisée, est transposée à celle de la *forme* des documents (au sens de description formelle, par exemple un ensemble de mots-clefs) avec une requête *formelle*. Le premier écart mis en évidence dans la définition est dû à l'interprétation du document nécessaire à sa description. Le second a pour cause la « traduction » du besoin d'information en une requête formelle, ce qui fait dire à C. J. Van Rijsbergen que seul l'utilisateur est en mesure de dire si un document contient l'information qu'il cherche, et que pour une même requête posée par deux utilisateurs différents un document pourra être pertinent pour l'un et non-pertinent pour l'autre.

Ainsi, il apparaît que le pari en recherche d'information consiste à réduire l'écart entre le niveau du *contenu des informations* et celui de la *forme des descriptions* (tant du côté des requêtes que de l'indexation), c'est à dire entre le niveau du *signifié* et

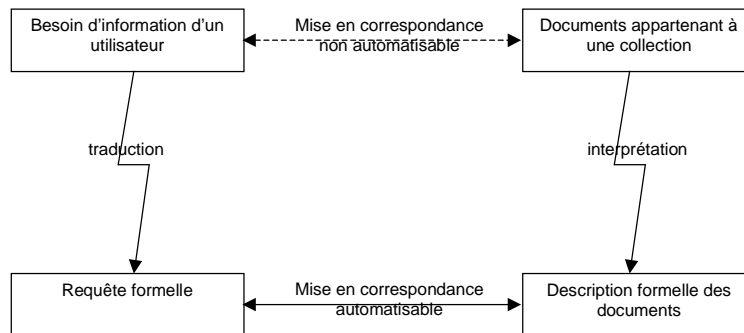


Figure 1 – Correspondances et écarts en recherche d'information

celui du *signifiant*. La recherche d'information est donc avant tout un problème de *sémantique*.

Nous choisirons tout d'abord une des théories linguistiques de la sémantique comme cadre à notre travail. Ensuite, nous exposerons notre modèle de recherche de documents incluant tant les aspects automatiques qu'interactifs. Enfin, nous généraliserons notre modèle à certaines tâches liées à la recherche de documents comme l'indexation, l'annotation ou la collaboration.

2. THEORIES DE LA SEMANTIQUE ET RECHERCHE D'INFORMATION

Dans l'esprit de l'évolution observée en informatique tendant à remplacer le paradigme du « tout automatique » par celui de « l'interactif », nous nous intéresserons à une théorie linguistique de la sémantique de tradition non plus logique mais rhétorique : la *sémantique interprétative* de François Rastier [2].

Ce que nous retenons de cette théorie pour le problème qui nous occupe est tout d'abord que la sémantique ne serait pas tant une question de *représentation* que de *communication*. D'une part, ceci nous permet, comme Klaus Tochtermann [3], d'affirmer que l'on ne peut parler de « bibliothèques virtuelles » que si l'on offre en plus des collections de documents numériques des « lieux » de communication et de collaboration. En effet dans une bibliothèque « traditionnelle », les bibliographies, dossiers, index, fiches de lecture sont autant de traces (incomplètes) de la communication qui permet à l'utilisateur de profiter de l'expertise d'un bibliothécaire ou d'un autre usager. D'autre part, ceci nous permet de comprendre certaines spécificités humaines de l'indexation qui n'apparaissent alors plus comme des défauts d'objectivité mais comme le réceptacle même du sens. En effet, Claire David et son équipe [4] ont mis en évidence que les différences d'indexation obtenues par des indexeurs professionnels pouvaient s'expliquer par leur expertise elle-même. Cette expertise les amène à des choix stratégiques différents (prenant par exemple en compte l'utilisateur) pour obtenir une solution partielle à ce problème mal-défini qu'est l'indexation.

Un autre point capital selon nous concerne les rapports qu'entretiennent l'occurrence d'un signe et son type. C'est en effet à partir des relations entre types (cf. hiérarchies de type chez John F. Sowa [5]) qu'il devient possible de faire des inférences sur les occurrences. Or François Rastier précise que l'approche logique, issue de la tradition aristotélicienne, considère cette hiérarchie des types comme évidente et donnée *a priori*, à tel point qu'on l'appelle souvent *ontologie* (du grec *ontos* : « ce qui est »). A contre pied, il propose que les types soient considérés comme des reconstructions transitoires et que l'on préfère à l'opposition *type/occurrence* l'opposition *occurrence-source/reprise*, chaque reprise étant même une nouvelle interprétation de l'occurrence-source. Cette vision d'une structure dynamique de la connaissance nous paraît plus ajustée à un système destiné à des scientifiques (au

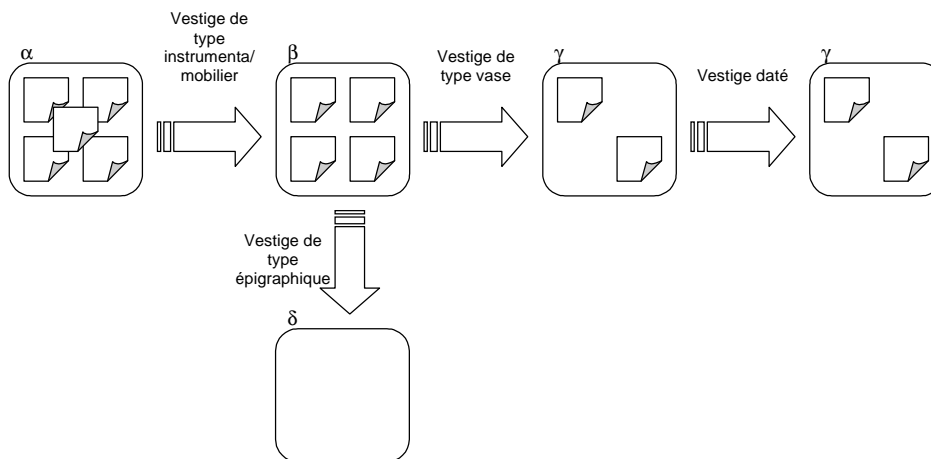


Figure 2 – La recherche de documents comme navigation entre corpus.

sens large). Les praticiens de la science moderne ne sont-ils pas appelés *chercheurs* plutôt que *savants* ?

Un dernier aspect que nous souhaitons souligner réside dans le fait que ces reconstructions transitoires sont réalisées « selon les objectifs de la pratique en cours ». Ceci nous amène à penser que les tâches d'indexation et de recherche de documents observées dans une bibliothèque universitaire ne peuvent prendre un sens que si on les intègre, à la manière de Kenton O'Hara et son équipe [6], dans une activité plus large. Ainsi, dans le cadre de notre étude qui vise à offrir à une communauté de chercheurs en archéologie un accès à un périodique spécialisé, la recherche d'information est à relier à l'ensemble des tâches (personnelles et interpersonnelles) exercées par les chercheurs en présence d'un tel corpus :

- la recherche d'un fragment de document en fonction d'un certain nombre de critères croisés,
- la recherche d'un fragment de document relié explicitement à un fragment donné (référéncé ou partageant une partie de sa structure logique),
- la recherche d'un fragment de document relié implicitement à un fragment donné (relation mise en évidence par leur présence commune dans une même bibliographie...),
- l'indexation officielle « pour le compte » d'une bibliothèque ou du périodique,
- la rédaction d'une fiche de lecture (résumé, extraits, étude critique...),
- la constitution d'une bibliographie sur un sujet donné,
- la communication d'une bibliographie et de son appareil critique (de manière publique dans un article ou privée entre un professeur et un étudiant),
- l'échange de références bibliographiques entre collègues.

3. MODELE POUR LA RECHERCHE DE DOCUMENTS

3.1 Processus de définition et interactivité

Le processus de recherche d'un document dans un corpus peut être identifié comme étant un processus de *définition* d'un document, *définition* au sens de *distinction*. Il s'agit en effet de distinguer un unique document parmi les documents du corpus. Ce processus de distinction d'un document par rapport au corpus de départ peut être décomposé en un ensemble de distinctions élémentaires qui appliquées successivement forment des chemins à travers des corpus de plus en plus petits. Prenons un exemple (fig.2), le corpus de départ contient cinq documents. Après avoir établi comme contrainte « Vestige de type instrumenta/mobilier », l'utilisateur obtient un corpus ne contenant plus que quatre documents. A partir de ce nouveau corpus,

si l'utilisateur pose comme contrainte « vestige de type épigraphique », il obtient un corpus vide, par contre s'il pose la contrainte « vestige de type vase », il obtient un corpus de deux documents. Enfin, à partir de ce corpus de deux documents, par la pose de la contrainte « vestige daté », l'utilisateur obtiendra le même corpus de deux documents.

Ainsi, on peut considérer la définition de documents comme une navigation entre corpus : l'application de certaines contraintes se traduisant par un « saut » vers un corpus plus spécialisé. L'apport de l'outil informatique doit être de guider l'expert dans sa navigation entre corpus. Le lecteur notera que nous limitons la mission de l'ordinateur à ce qui est purement *formel* et laissons à la charge de l'être humain ce qui concerne la *substance* des choses. C'est parce que nous sommes convaincus que l'intelligence sera toujours du côté de l'humain et non de l'automate qu'il nous semble indispensable de mettre l'interactivité au cœur de la recherche de documents.

3.2 Apport de l'outil informatique

Tâchons maintenant d'identifier un apport possible de l'outil informatique dans cette recherche de documents définie comme navigation entre corpus. Reprenons le même exemple (fig.2), on remarque que pour un corpus donné :

- certaines contraintes mènent au corpus vide : on parlera de contraintes *impossibles* (ex : aucun document du corpus β ne satisfait à la contrainte « vestige de type épigraphique »),
- certaines contraintes «bouclent » sur un corpus : on parlera de contraintes *connues* (ex : tous les document du corpus γ satisfont à la contrainte « vestige daté »),
- pour les autres contraintes, on parlera de contraintes *possibles* (ex : certains documents du corpus β satisfont à la contrainte « vestige de type vase »).

A la lumière de ces trois définitions, on peut construire, un système de « navigation » entre corpus permettant d'assister l'utilisateur dans le raffinement de ses requêtes. Néanmoins sa mise en place à partir d'un simple index de type documents/mots-clés, pose problème dans le contexte qui nous intéresse. En effet, pour peu que le vocabulaire du domaine scientifique concerné soit un peu riche, le nombre de mots-clés candidats pour l'indexation et la recherche risque de dépasser tant les capacités cognitives de l'utilisateur que calculatoires de l'ordinateur (dans un contexte interactif et dynamique). Aussi, nous proposons d'adopter comme descrip-

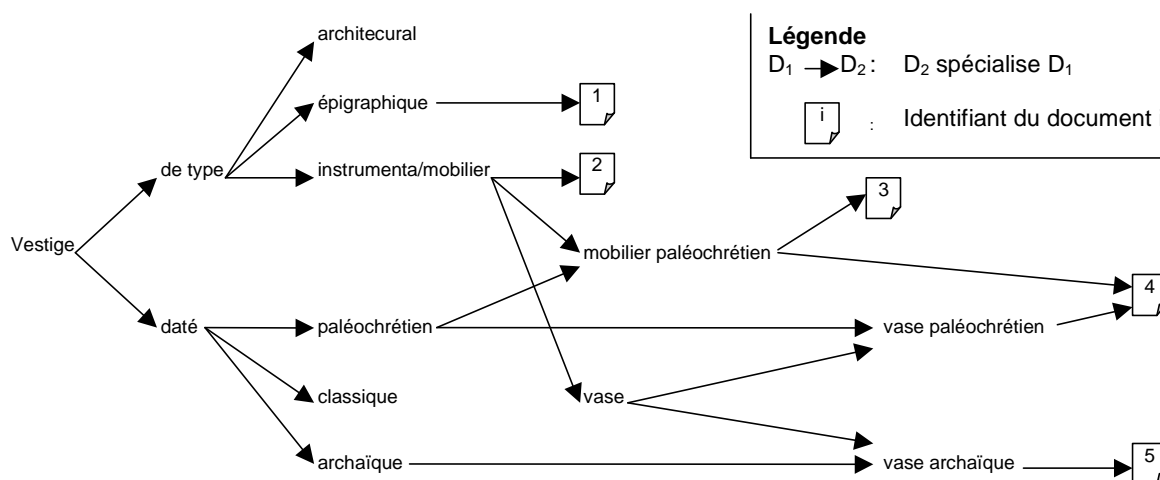


Figure 3 – Graphe acyclique indexant cinq documents traitant de fouilles funéraires.

tion des documents, non plus des descripteurs isolés, mais des descripteurs organisés par un ordre partiel (fig.3) conformément à des relations de spécialisation¹. Il devient alors possible de dire par exemple que « mobilier paléochrétien » et « vase » ont pour généralisation commune « instrumenta/mobilier », ou encore que « vase » et « paléochrétien » ont pour spécialisation commune « vase paléochrétien »².

Par cette structuration des descripteurs, il ne s'agit plus, pour un état donné, de donner une valeur {connue, possible, impossible} à tous les descripteurs mais uniquement aux « enfants » du descripteur « racine » et aux enfants de ses enfants lorsque ces derniers sont « connus » (et ainsi de suite). La navigation entre corpus (fig.4) s'effectue alors à partir de la sélection du descripteur racine et par sélection successive de descripteurs « possibles » ou « désélection » de descripteurs précédemment sélectionnés. Suivons pas à pas le scénario présenté par la Figure 4 :

Etape 1. Le corpus global traite de « *vestiges typés* ». Les corpus plus spécialisés traitent de *vestiges datés*, de *vestiges de type épigraphique* ou de *vestiges de type instrumenta/mobilier* mais pas de *vestiges de type architectural* (cette description ne correspond en effet à aucun document du corpus considéré). Après sélection par l'utilisateur de *instrumenta/mobilier*, le système passe à l'étape 2.

Etape 2. Le corpus sélectionné traite de *vestiges de type instrumenta/mobilier*. Cette description correspond exactement au document ayant 2 pour identifiant. Les corpus plus spécialisés traitent de *mobilier paléochrétien* ou de *vases*. Aucun ne traite de *vestiges de type architectural* ou *épigraphique*. Après sélection par l'utilisateur de *vase*, le système passe à l'étape 3.

Etape 3. Le corpus sélectionné traite de *vestiges de type vase (instrumenta/mobilier)* et de *vestiges datés*. On remarque que le fait qu'ils soient *datés* est inféré par l'ordinateur (tous les documents du corpus traitant de *vases* traitent de *vestiges datés*). Cette description correspond exactement au document ayant 2 pour identifiant. Les corpus plus spécialisés traitent de *mobilier paléochrétien*, de *vases paléochrétiens*, de *vases archaïques*, de *vestiges paléochrétiens* ou de *vestiges archaïques*. Aucun ne traite de *vestiges de type architectural* ou *épigraphique*, ni de *vestiges datés de l'époque classique*. Après sélection par l'utilisateur d'*archaïque*, le système passe à l'étape 4.

Etape 4. Le corpus sélectionné traite de *vestiges de type vase (instrumenta/mobilier) datés de l'époque archaïque*. Ce corpus ne contient qu'un seul document : celui ayant 5 comme identifiant.

4. APPLICATION AUX AUTRES TACHES DOCUMENTAIRES

Tâchons maintenant d'étudier comment le modèle que nous venons de définir pour la recherche de documents peut s'appliquer également à l'indexation et l'annotation (coopérative) de documents (structurés ou non).

4.1 Indexation de documents

Tout comme la recherche de documents, le processus d'indexation d'un document dans un corpus peut être identifié comme étant un processus de *définition* d'un document. Notre système de navigation parmi les corpus, permet à l'indexeur de se déplacer rapidement vers un corpus qui pourrait contenir le document à indexer mais dont les sous-corpus actuels ne le pourraient pas.

¹ Cette spécialisation est à prendre au sens que si un descripteur D_2 spécialise un descripteur D_1 , tout document décrit par D_2 l'est aussi par D_1 .

² Bien distinguer la composition de descripteurs, de leur simple cooccurrence : un document peut traiter d'un vestige « paléochrétien » et d'un « vase », sans pour autant traiter d'un « vase paléochrétien ».

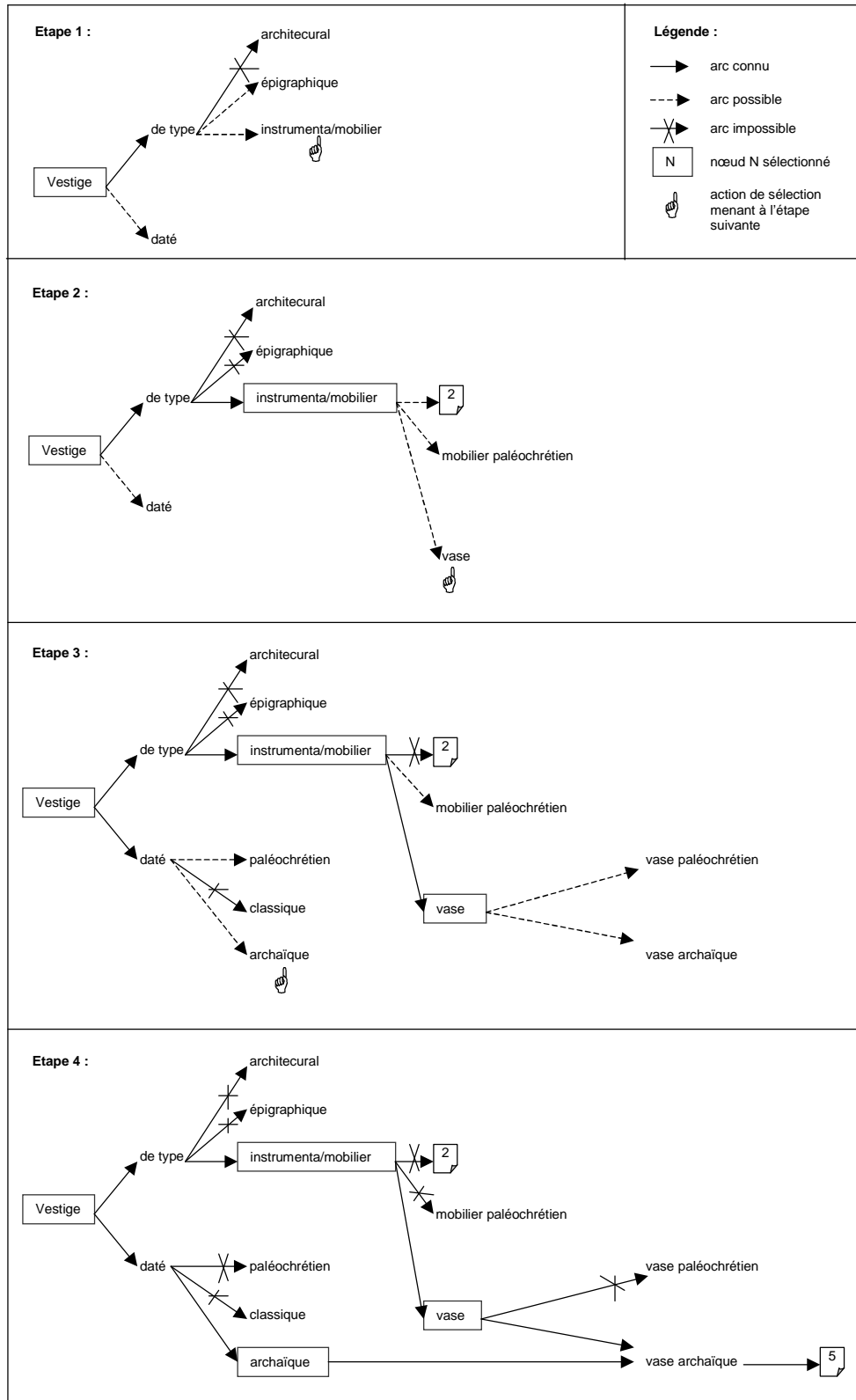


Figure 4 – Scénario en quatre étapes d'une navigation dans un graphe d'indexation (fig.3).

Par exemple, pour ajouter dans l'index de la figure 3 un document traitant de *vases classiques*, on pourra suivre les étapes 1 à 3 de la figure 4. En effet, le document à indexer pourrait appartenir au corpus traitant des vases mais pas à ceux traitant des *vases paléochrétiens* ou *archaïques*.

L'indexeur peut alors rattacher l'identifiant de son document aux descripteurs connus, à certains descripteurs possibles, ou encore à des spécialisations de descripteurs connus qu'il créera pour l'occasion. L'opération peut se répéter afin d'associer l'identifiant du document ou les nouveaux descripteurs à d'autres corpus. Ainsi, toujours dans notre exemple, l'indexeur pourra rattacher un nouveau descripteur *vase classique* à *vase*, puis rattacher l'identifiant du document à *vase classique*, et enfin rattacher *vase classique* à *classique*.

Utiliser l'algorithme de navigation entre corpus non seulement pour la recherche mais aussi pour l'indexation, nous permet de faciliter la réutilisation des index précédents. Il s'agit en effet de favoriser les recoupements sans pour autant imposer un « vocabulaire contrôlé ».

4.2 Cas particulier des documents structurés

Nous proposons maintenant d'étendre aux documents structurés les notions introduites ci-dessus. Pour Franck Fourel [7], la structure d'un document est organisée selon trois types de relations : celles de composition (ex : un chapitre est composé de sous-chapitres), celles de séquence (ex : un paragraphe suit un autre) et celles de référence (ex : une phrase fait référence à une figure). La question est alors de savoir comment se propage l'attribut³ d'un élément structurel⁴ suivant ces relations. Pour reprendre son exemple, doit-on dire qu'un document est à l'état de brouillon lorsqu'au moins un chapitre est à l'état de brouillon, lorsqu'ils le sont tous, lorsqu'un certain pourcentage l'est ou enfin est-il impossible de déduire l'état du document à partir de l'état des chapitres ?

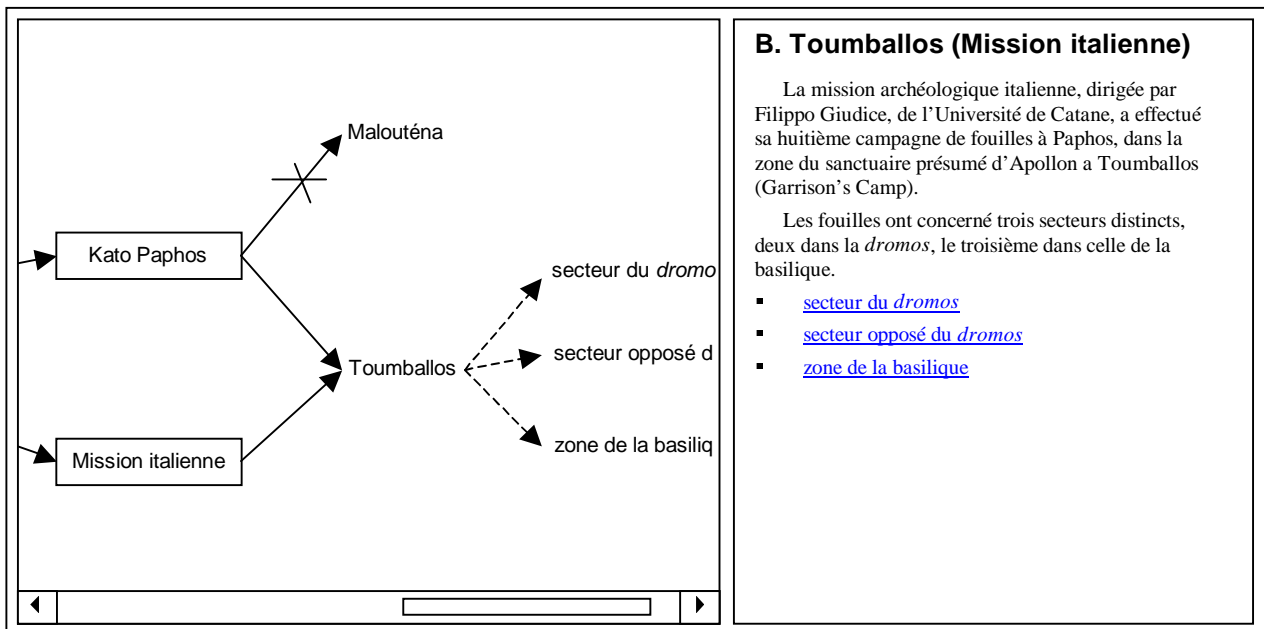


Figure 5 – Navigation conjointe dans le graphe d'indexation et le document structuré.

³ ex : « auteur = Tartempion », « date_modification = 04/02/2000 »

⁴ ex : paragraphe, figure, chapitre, document...

Tout d'abord, nous plaçant dans un contexte beaucoup plus limité, nous ne considérerons qu'un type de relation. En effet, d'une part, nous négligerons les relations de séquence qui ne nous paraissent pas indispensables pour la recherche proprement dite, et d'autre part nous regrouperons dans le même concept de « référence acyclique »⁵ la composition et la référence. Nous serons donc amenés, dans le cas de références cycliques entre éléments structurels à considérer un élément virtuel faisant une référence acyclique à ces éléments.⁶ Ensuite, dans un souci de simplicité, nous poserons que la description du tout est applicable à la partie et que la description de la partie n'est pas applicable au tout. Ainsi, il devient possible d'utiliser notre système non plus uniquement avec des documents mais aussi avec des parties de documents. Il est à noter que dans notre système, une même partie de document peut appartenir à la fois à plusieurs structures différentes. L'interface homme-machine est telle que tous les documents correspondant à une contrainte « connue » ou « possible » – cf. définitions (§ 4.1.b) et (§ 4.1.c) seront affichés automatiquement (ou sous forme iconique). Il devient alors possible d'afficher un index, une carte, le document entier, un titre, une introduction ou tout autre fragment de document multimédia pouvant être considéré comme une étape dans la navigation vers une partie plus spécialisée. Précisons que la navigation pourra se faire conjointement à l'aide du graphe d'indexation ou des hyperliens des documents affichés.

Par exemple (fig.5), un fragment de document a été affiché à la suite d'une recherche multicritère dans le graphe d'indexation (le lieu *Kato Paphos* et l'entreprise de fouille *Mission italienne* ont été sélectionnés). D'autres fragments de documents, correspondant à des secteurs géographiques plus précis (*secteur du dromos*, etc.), peuvent être atteints soit à l'aide des hyperliens contenus dans le fragment de document soit par sélection des descripteurs correspondants.

4.3 Annotation et collaboration

Une annotation dans notre contexte est l'activité d'un utilisateur consistant à relier (par une relation de généralisation) un descripteur ou un identifiant (par exemple celui d'un mot dans un document) à un autre descripteur. Il est à noter que le des-

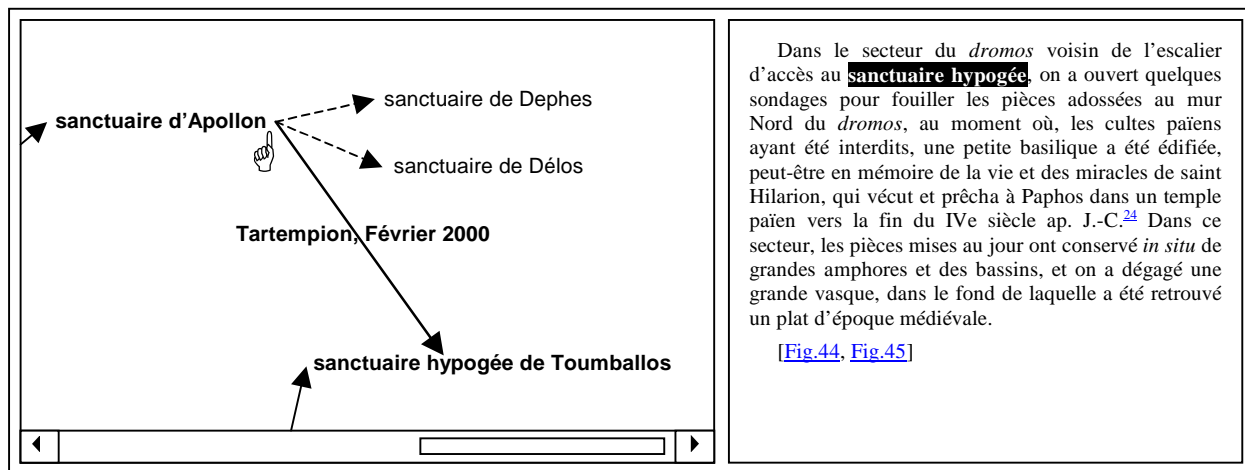


Figure 6 – Annotation dans un document par un lecteur¹.

⁵ Les références acycliques sont définies par opposition aux références cycliques pour lesquelles il existe un chemin dont le départ et l'arrivée sont le même élément.

⁶ ex : on pourra, par exemple, remplacer le fait que l'élément structurel traitant de l'*Artémision d'Ephèse* et celui traitant du *sanctuaire de Brauron* se fassent mutuelle-

cripteur ou l'identifiant annoté aura pu être créé par ce même utilisateur, par un autre ou par l'indexeur officiel.

Dans notre exemple (fig.6), Monsieur Tartempion, après avoir souligné *sanctuaire hypogée* dans le texte, et l'avoir identifié par *sanctuaire hypogée de Toumballos*, a rattaché (en février 2000) cet identifiant au descripteur préexistant *sanctuaire d'Apollon*.

Martin Röscheisen et son équipe [8] se sont intéressés à la question des annotations dans le contexte d'une communauté « en-ligne ». Dans leur modèle, sont gérées les informations portant sur la création des annotations, ainsi que les droits de lecture et de modification. La répartition des annotations sur le réseau est fonction des droits qui y sont attachés, afin d'obtenir des serveurs spécialisés pour des groupes d'utilisateurs.

De même, dans notre système, chaque arc du graphe est étiqueté par son auteur et sa date de création. D'autre part, il est possible, de définir des accès publics (par exemple pour des travaux publiés), privés (pour des travaux non-publiés) ou réservés à un groupe (par exemple pour les étudiants d'un cours ou les membres d'une équipe de recherche).

En ce qui concerne la répartition des données et leur réplication, nous envisageons de généraliser l'approche en l'appliquant non seulement aux droits d'accès mais aussi à la consultation effective des annotations ainsi qu'à la notion de proximité⁷ des utilisateurs. Il s'agirait en effet d'utiliser pleinement la métaphore de la diffusion des savoirs au sein de la communauté scientifique.⁸

5. BILAN ET PERSPECTIVES

Nous avons donc étudié une application possible dans le domaine de la recherche de documents de la théorie linguistique de la « sémantique interprétative ». Nous avons été amené à identifier la recherche de documents comme une navigation entre corpus et à définir ce qui relevait de l'interactivité et ce qui relevait du traitement informatique. Ensuite, nous avons intégré la recherche de documents dans une activité plus large, plus complexe, impliquant davantage l'utilisateur et comprenant entre autres l'enrichissement de la description des documents (simples ou structurés) et le partage de cette description avec d'autres utilisateurs.

Le travail présenté ici s'inscrit dans le cadre du projet de « mise en ligne » d'une chronique de fouilles archéologiques [9] [10], projet lié à une thèse en cours sur la modélisation de la documentation de consultation pour experts, sous la tutelle d'un laboratoire de recherche en informatique et d'une école doctorale en archéologie. A l'heure actuelle, nous disposons d'un prototype en Java (fig.7) qui nous a permis de mettre en évidence la validité de l'algorithme de filtre de graphes acycliques ainsi que la faisabilité d'une interface homme-machine basée dessus.

Dans les mois à venir, nous prévoyons d'éprouver notre prototype, tant au niveau algorithmique avec de gros volumes de données, qu'au niveau interface auprès des utilisateurs (pour cela nous devons faire évoluer le prototype vers une version client-serveur).

Ainsi, dans les prochaines années, nous pourrions offrir aux chercheurs un système leur permettant de prendre des « notes de lecture », de les partager, et de « naviguer » à travers le réseau qu'elles constituent ; le pari étant, tel que le souhaitait Vannevar Bush [11], d'assister le lecteur de publications savantes en le soula-

ment référence par le fait qu'un nouvel élément structurel traitant des *sanctuaires d'Artémis* leur fasse référence.

⁷ On se rapproche ici de la notion de « proxy ».

⁸ Les interactions entre chercheurs sont dépendantes des thèmes de recherche mais aussi de leur proximité géographique ou institutionnelle.

geant des aspects répétitifs de son activité et en le laissant se concentrer sur les aspects créatifs, intuitifs et à haut niveau d'abstraction.

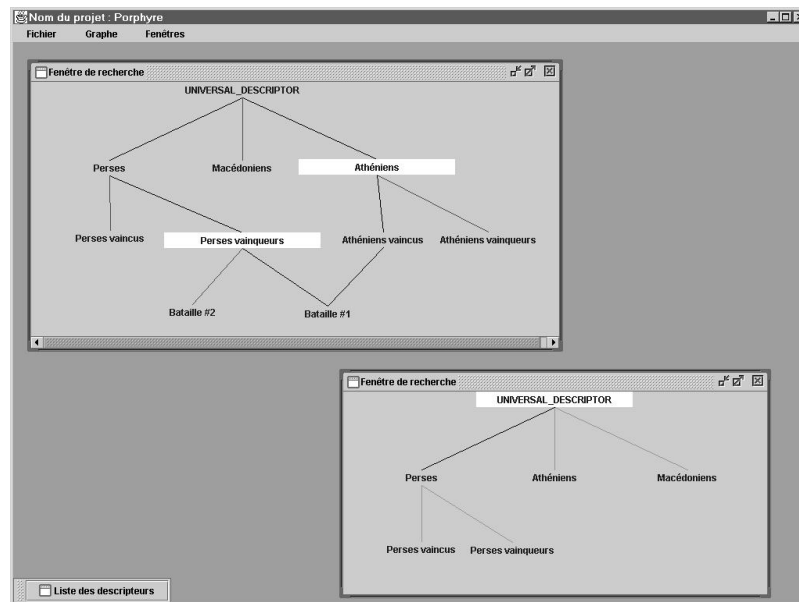


Figure 7 – Copie d'écran du prototype en Java du module de navigation.

6. REFERENCES BIBLIOGRAPHIQUES

- [1] **VAN RIJSBERGEN C. J.**, "A new theoretical framework for information retrieval", *Proceedings of 1986 ACM Conference on Research and Development in Information Retrieval*, 1986.
- [2] **RASTIER F.**, "Sens et signification". *Protée*, Printemps 1998 ; version augmentée parue dans *l'Encyclopédie philosophique universelle*, PUF, 1999.
- [3] **TOCHTERMANN K.**, "A first step toward communication in virtual libraries", *Digital Libraries '94: Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries*, 1994.
- [4] **DAVID C., GIROUX L., BERTRAND-GASTALDY S. et LANTEIGNE D.**, "Indexing as Problem Solving: a Cognitive Approach to Consistency", *ACSI'95: Annual Conference of the Canadian Association for Information Science*, 1995.
- [5] **SOWA J. F.**, "Semantic Networks", *Encyclopaedia of Artificial Intelligence*, Wiley, New-York, 1987 ; revised edition, 1992.
- [6] **O'HARA K., SMITH F., NEWMAN W. et SELLEN A.**, "Student readers' use of library documents: Implications for library technologies", *CHI'98 Conference proceedings on Human factors in computing systems*, 1998.
- [7] **FOUREL F.**, *Modélisation, indexation et recherche de documents structurés*, chapitre 6 : Indexation structurelle, Thèse de doctorat en informatique, Université de Grenoble 1, 1998.
- [8] **RÖSCHEISEN M., MOGENSEN C. et WINOGRAD T.**, "Beyond browsing: Shared comments, soaps, trails, and on-line communities", *The Third International World Wide Web Conference, "Technology, Tools and Applications"*, 1995.
- [9] **BÉNEL A., CALABRETTO S., et PINON J. M.**, "Indexation "sémantique" de documents archéologiques". *Actes du deuxième colloque du chapitre français de l'ISKO, "L'indexation à l'heure d'Internet"*, 1999.
- [10] **BÉNEL A., CALABRETTO S., PINON J. M., et IACOVELLA A.**, "Vers un outil documentaire unifié pour les chercheurs en archéologie", *Actes du XVIIIe congrès INFORSID*, 2000.
- [11] **BUSH V.**, "As we may think". *The Atlantic Monthly*. July 1945.