

## Chapitre 7. Réseaux de description<sup>\*</sup>

*The real heart of the matter of selection, however, goes deeper than a lag in the adoption of mechanisms by libraries or a lack of development of devices for their use. Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing. When data of any sort are placed in storage, they are filed alphabetically or numerically, and information is found (when it is) by tracing it down from subclass to subclass. It can be in only one place, unless duplicates are used; one has to have rules as to which path will locate it, and the rules are cumbersome.*  
Vannevar Bush, *As we may think*, 6.

Dans le chapitre précédent, nous avons défini la notion d'objet documentaire. Dans ce chapitre-ci, nous verrons comment en décrivant de manière structurée ces objets, nous les mettrons en contexte les uns avec les autres. Suivant la nature des objets documentaires, le résultat de cette mise en contexte pourra être interprété comme un document (au sens traditionnel), un corpus de documents, ou encore comme un document « sur mesure ». Autrement dit, nous ne ferons pas de différence formelle entre :

- la description d'une section parmi un document (rédaction),
- celle d'un document parmi une collection (indexation),
- ou encore celle d'un fragment parmi un dossier bibliographique (annotation).

Dans un premier temps, nous étudierons comment la question de la description de documents (de manière structurée) est ordinairement traitée. Dans un deuxième temps, nous proposerons un modèle basé sur la notion de « point de vue ». Enfin, dans un dernier temps, nous verrons comment il est possible d'offrir des outils basés sur notre modèle permettant « d'arpenter » l'espace documentaire.

---

<sup>\*</sup> Des parties de ce chapitre ont fait l'objet des conférences suivantes :

- l'International Symposium on Methodologies for Intelligent Systems [BenelEtAl02],
- le Colloque International Francophone sur l'Écrit et le Document [BenelEtAl00b],
- le Congrès INFORSID [BenelEtAl00a],

## 1. Décrire de manière structurée des objets documentaires

### a. Des arbres qui cachent... la bibliothèque

L'exergue de ce chapitre rappelle que l'organisation traditionnelle des bibliothèques est basée sur l'idée que chaque livre traiterait d'un sujet unique, sujet qui lui-même serait situé sans ambiguïté possible dans une hiérarchie universelle. L'auteur (Vannevar Bush) dénonce ici le caractère artificiel et inutilisable de ce modèle arborescent instauré dans les bibliothèques par Dewey (1859-1952).

Les racines du problème remontent sans doute à Aristote (IV<sup>e</sup> s. av. J.-C.). Celui-ci nous déclare en effet dans les *Catégories* :

Quand une chose est attribuée à une autre comme à son sujet, tout ce qui est affirmé du prédicat devra être aussi affirmé du sujet [...].

Si les genres sont différents et non subordonnés les uns aux autres, leurs différences seront elles-mêmes autres spécifiquement [...].

Par contre, dans les genres subordonnés les uns aux autres, rien n'empêche que leurs différences soient les mêmes, car les genres plus élevés sont prédicats des genres moins élevés, de sorte que toutes les différences du prédicat seront aussi des différences du sujet. [Aristote-300, *Catégories* 3]

Ces trois principes sont interprétés par Porphyre<sup>65</sup> (III-IV<sup>e</sup> s. ap. J.-C.) comme définissant un arbre dans lequel à chaque nœud un genre (*genus*) se divise en sous-genres en fonction d'une différence (*differentia*). Chacun de ces trois principes peut donc se lire de manière appliquée dans « l'arbre de Porphyre » (cf. Figure 7.1) :

1. Les hommes sont des animaux. Les animaux sont sensibles. Donc les hommes sont sensibles.
2. « Rationnel » est une propriété « d'Homme » qui ne subordonne pas « Végétal ». Donc, « Rationnel » n'est pas une propriété de « Végétal ».
3. L'Homme est :
  - un animal rationnel,
  - un être vivant sensible et rationnel,
  - une matière animée, sensible et rationnelle,
  - une substance corporelle, animée, sensible et rationnelle.

---

- l'Escuela interlatina de altos estudios en linguística aplicada [BenelEtCalabretto00],

- le Chapitre français de l'ISKO [BenelEtAl99].

<sup>65</sup> dans son introduction aux *Catégories*.

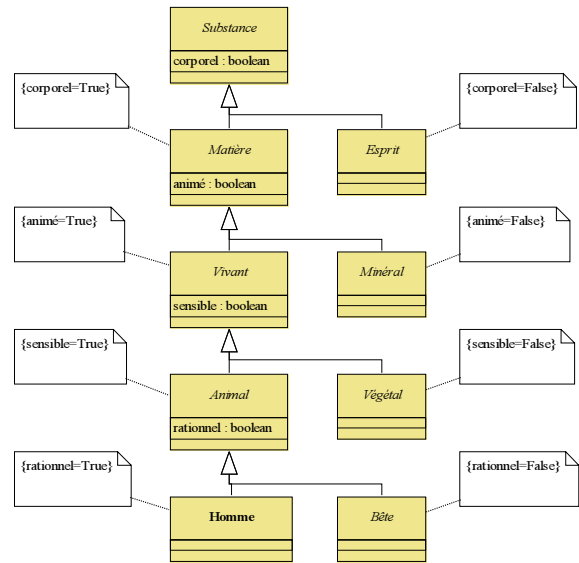


Figure 7.1 : « L’arbre de Porphyre » : Fresque du XVIII<sup>e</sup> s.<sup>66</sup> et diagramme de classe UML.

On comprend sans peine qu’une organisation des connaissances d’une telle esthétique et d’une telle efficacité soit devenue le parangon de la pensée occidentale. Cependant, si cette méthode est tout à fait valide pour parler de classes, elle ne devrait en aucun cas être utilisée pour des instances, celles-ci pouvant souvent être placées dans plusieurs classes contradictoires. En effet, on pense tout de suite au célèbre exemple de Nixon potentiellement pacifiste en tant que quaker et belliciste en tant que républicain.

### b. Alternatives

La méthode arborescente étant inutilisable pour classer des instances, *a fortiori* elle l’est également pour des livres dont la description pourra contenir des classes<sup>67</sup>, des

<sup>66</sup> Détail d’une fresque du monastère de Schussenried (Allemagne) ; Artiste : Hermann, 1757 ; Photographie : J. Garrett, 2000 ; Disponible sur Internet : <<http://www.library.northwestern.edu/collections/garrett>>

<sup>67</sup> Ou tout au moins des objets génériques.

instances et des liens entre instances. Aussi, depuis longtemps, des alternatives au modèle d'indexation hiérarchique ont été proposées<sup>68</sup>. La plus connue en sciences de l'information est celle de Ranganathan (1872-1972) appelée aussi « indexation par facettes », mais les plus ambitieuses sont sans doute celles basés sur les graphes conceptuels de John F. Sowa [MechkourEtAl95, Martin96, Genest00]. Dans ces dernières, chaque document est décrit par un graphe (distinct) comprenant des objets (éventuellement génériques) et des liens entre ces objets (cf. Figure 7.2). Ces objets et ces liens, sont des instances d'un modèle du domaine<sup>69</sup> (cf. Figure 7.3).

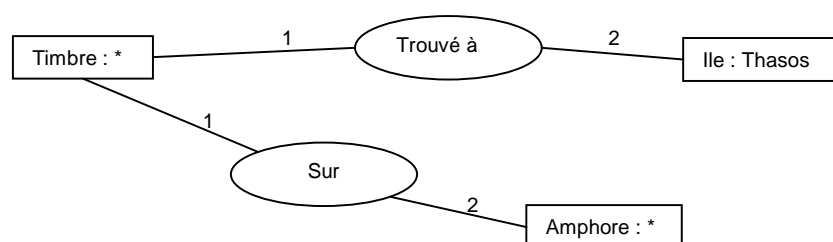


Figure 7.2 : Indexation à l'aide des graphes conceptuels d'une monographie traitant des timbres amphoriques thasiens.

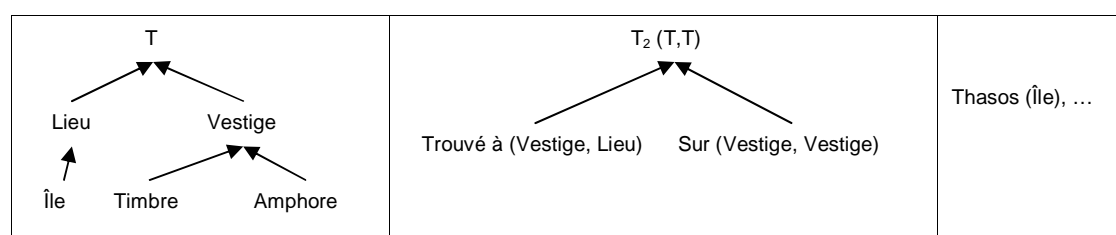


Figure 7.3 : Modèle du domaine nécessaire à l'indexation de la figure précédente : concepts, relations et instances.

Malheureusement, malgré leur puissance d'expression, ces alternatives sont, comme nous allons le voir, assez peu adaptées au cas qui nous occupe : celui d'une modélisation *dynamique* effectuée *par les usagers* de la bibliothèque.

<sup>68</sup> Notons que la même critique a été faite pour l'organisation des objets documentaires dans un document, en réaction à la suprématie du format XML dans ce domaine. Voir par exemple [RenearEtAl96] et [Nelson97] ou encore le modèle proposé par notre groupe de réflexion au sein de l'ISDN pour les documents à structure multiple.

<sup>69</sup> Notons que cette séparation entre le modèle du domaine et les graphes d'indexation reprend celle existant dans les bibliothèques entre les « autorités » et les fiches d'indexation.

**c. Réfutation**

Prenons un exemple en archéologie. Philippe Bruneau [Bruneau76], en réaction aux premières « banques de données archéologiques », faisait remarquer l'impossibilité de décrire « objectivement » une photographie du type de la Figure 7.4. Était-on en présence de la représentation d'une mosaïque noire sur fond blanc ou blanche sur fond noir ? Plus grave encore, l'auteur nous faisait même douter du bien fondé d'une telle typologie.



Figure 7.4 : Mosaïque noire sur fond blanc ou blanche sur fond noir ? [Bruneau76]

Dans un tel cas, nous devons disposer d'un modèle permettant d'exprimer qu'un premier point de vue affirme qu'il s'agit d'une mosaïque noire sur blanc, qu'un second affirme l'inverse, et qu'un troisième propose une typologie toute autre.

Les deux premiers points de vue étant contradictoires, notre « modèle de connaissance » devra être beaucoup plus permissif que la normale :

- Les structures seront non hiérarchiques (graphes orientés acycliques),
- Il n'y aura pas de négation (donc pas de principe de tiers-exclu, ni de principe de non-contradiction),
- Les points de vue ne seront pas dépendants les uns des autres, si ce n'est par l'intermédiaire des corpus décrits.

Du fait que le troisième des points de vue remette en cause la typologie utilisée dans les deux premiers, nous ne pourrons plus considérer qu'il existe *un* modèle *fixe* du domaine, mais plutôt *des* modèles *hypothétiques* et *transitoires*, évoluant de pair avec leurs instances. La séparation des classes et des instances en deux espaces apparaît par

conséquent inutile. De manière plus générale, l'aspect dynamique de la modélisation empêchera un typage trop fort des primitives<sup>70</sup>.

Les descriptions n'étant plus normées, il sera impossible de connaître *a priori* leur forme. Les interactions homme-machine ne devront donc pas suivre le modèle question-réponse mais plutôt celui de la navigation. La recherche de documents se fera donc de manière itérative et ira de pair avec une découverte de la structure du corpus. Dans une telle approche, la description du document sera un sous-graphe de la description du corpus<sup>71</sup>. En ce sens, nous nous rapprocherons un peu des techniques qui visent à agréger des graphes disjoints afin de donner une vision d'ensemble [Chalendar97, PredigerEtWille99, BurrowEtEklund94, EklundEtCole02].

Enfin, le fait que les experts ne soient pas des professionnels de la modélisation, nous encourage à proposer un modèle dont l'utilisation pour des descriptions simples sera assez intuitive, et dans lequel, il sera possible, moyennant une formation, d'établir des descriptions plus précises. Nous donnerons plus loin une esquisse du contenu d'une telle formation (cf. Chapitre 11).

## 2. Un modèle à base de points de vue

Notre modèle appelé *réseau de description* se présente sous la forme d'un graphe orienté acyclique (cf. Figure 7.5). Les nœuds sont appelés des « *descripteurs* » et les arcs des « *spécialisations* ». Un arc orienté entre les descripteurs *A* et *B* se lit : « tout objet documentaire décrit par *B* l'est aussi par *A* ».

---

<sup>70</sup> Que l'informaticien qui n'a jamais changé, au cours d'un processus de modélisation, une instance en classe, ou une classe concrète en classe abstraite, nous jette la première pierre...

<sup>71</sup> Contrairement aux travaux cités précédemment où chaque document était décrit par un graphe distinct des autres.

## CHAPITRE 7. RÉSEAUX DE DESCRIPTION\*

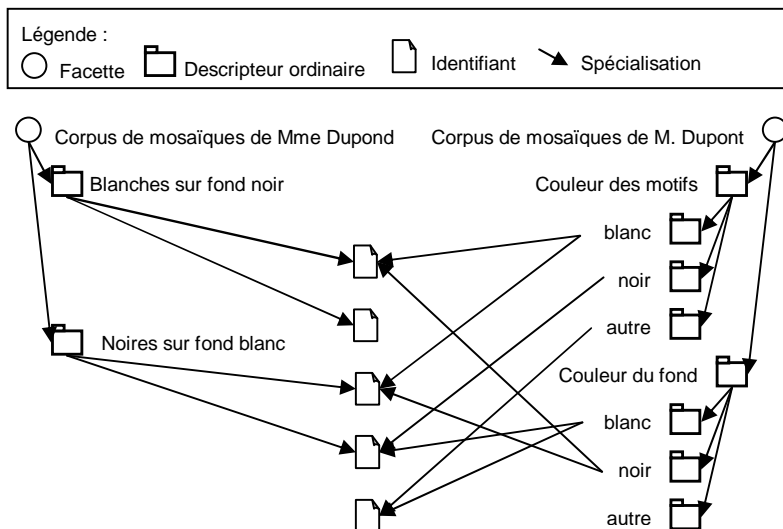


Figure 7.5 : Exemple de réseau de description

Il est important de mentionner que seul les nœuds et les arcs sont utiles pour le système. Mais de sorte que les usagers puissent interpréter le réseau, nous associons à chaque nœud une étiquette et à chaque arc son historique (cf. Chapitre 5). Du moment que la définition formelle des arcs est respectée, l'utilisateur est libre d'utiliser ces arcs pour modéliser (cf. Chapitre 11) des taxinomies, des méronymies, des instanciations...

Parmi les descripteurs (cf. Figure 7.6), certains ne sont pas généralisables : on les appelle des « facettes ». Chacun correspond à un point de vue indépendant. D'autres ne sont pas spécialisables, on les appelle des « *identifiants* ». Chacun fait référence à un objet documentaire unique.

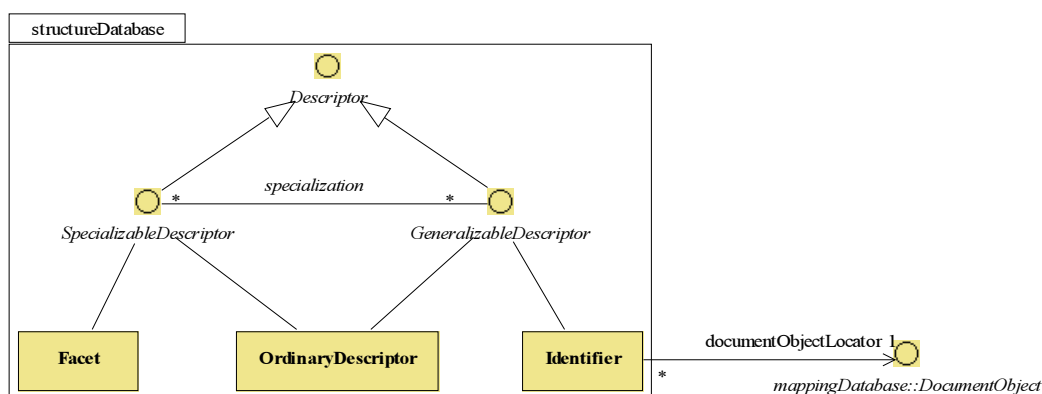


Figure 7.6 : Réseau de description (diagramme de classe UML)

Plus formellement, nous avons affaire aux ensembles suivants : *Descriptor*, *SpecializableDescriptor*, *GeneralizableDescriptor*, *Identifieur*, *Facet*, *DocumentObject*.

Ces ensembles satisfont axiomatiquement les relations suivantes :

$$Descriptor = SpecializableDescriptor \cup GeneralizableDescriptor$$

$$Identifieur \subset GeneralizableDescriptor$$

$$Facet \subset SpecializableDescriptor$$

Nous allons maintenant définir par des spécifications algébriques<sup>72</sup> :

- le schéma des données à stocker (primitives),
- les contraintes supplémentaires que ces données doivent respecter (contraintes),
- les requêtes complexes qui seront effectués sur ces données (définitions).

**Primitive** :  $specialization(\_, \_) : SpecializableDescriptor \times GeneralizableDescriptor \rightarrow Boolean$

**Primitive** :  $\_.getDOI : Identifieur \rightarrow DocumentObject$

Fonction permettant d'obtenir le document correspondant à un identifiant donné.

Note: le fait que ce soit une fonction et non une relation traduit une contrainte d'intégrité fonctionnelle sur les données.

**Définition** :  $describes(\_, \_) : Descriptor \times GeneralizableDescriptor \rightarrow Boolean$

Relation construite de sorte qu'elle soit réflexive et qu'elle constitue la fermeture transitive de la relation « *spécialization* ».

$$describes(x, x)$$

$$describes(x, y) \leftarrow specialization(z, y) \wedge describes(x, z)$$

---

<sup>72</sup> Voir en annexe l'explication de la syntaxe utilisée.



Note : cet algorithme est ascendant, ce qui constitue une optimisation pour un usage « normal » du réseau de description (un descripteur ayant souvent des centaines d'enfants et juste quelques parents).

**Contrainte** : « Acyclicité »

Aucun cycle ne doit exister dans le réseau de description.

$$\perp \leftarrow specialization(x,y) \wedge describes(y,x)$$

**Contrainte** : « Enracinement »

Un descripteur ne doit pas appartenir à plusieurs facettes.

$$\perp \leftarrow f_1 \in Facet \wedge f_2 \in Facet \wedge f_1 \neq f_2 \wedge describes(f_1,x) \wedge describes(f_2,x)$$

### 3. Arpenter l'espace documentaire

Chaque structure créée par un expert est perçue comme la trace d'un *point de vue*, d'une théorie portant sur une partie de la discipline. En effet, structurer son corpus documentaire, c'est structurer son objet d'étude, sa discipline.

Dès lors que des points de vue sont publiés (au sens de « rendus publics ») par des experts, tous ceux qui le souhaitent peuvent les voir superposés au corpus. Toutefois, si nous avons choisi de laisser chaque expert exprimer son point de vue, ce n'est pas pour glorifier une subjectivité débridée, mais pour autoriser le débat, pour viser l'intersubjectivité. Il s'agit donc de pouvoir comparer entre eux ces points de vue.

Le mécanisme que nous avons offert aux utilisateurs est un filtre de graphes. Il permet par induction totalisante de trouver des rapports entre descripteurs, non-dits au niveau des modèles, mais apparaissant dans leurs usages. Pour reprendre notre exemple de typologies de mosaïques, le système nous indiquerait que lorsque tel auteur décrit les mosaïques comme blanche sur fond noir, un autre les décrit « toujours » (ou « parfois », ou « jamais ») comme des mosaïques noires sur fond blanc.

Notons qu'en filtrant les informations, nous en réduisons la charge cognitive. Nous tenons donc compte de l'une des préoccupations centrale des systèmes interactifs de recherche d'information [Hearst99]. En effet, vu la taille importante que prendra le réseau de description au fur et à mesure de son utilisation, il est crucial de n'en présenter à un moment donné qu'une partie. En fournissant ainsi un moyen à l'utilisateur de s'orienter efficacement, nous espérons l'aider à arpenter l'espace documentaire.

### a. Filtrer des descripteurs « à plat »

Gerard Salton, à la fin des années soixante [Salton68], a défini un modèle pour la recherche d'information basé sur la théorie des ensembles. Ce modèle considère un ensemble des documents et un ensemble des « descripteurs ». Ainsi on peut tracer les graphes d'inclusion de corpus (cf. Figure 7.7) et de conjonction de requêtes (cf. Figure 7.8).

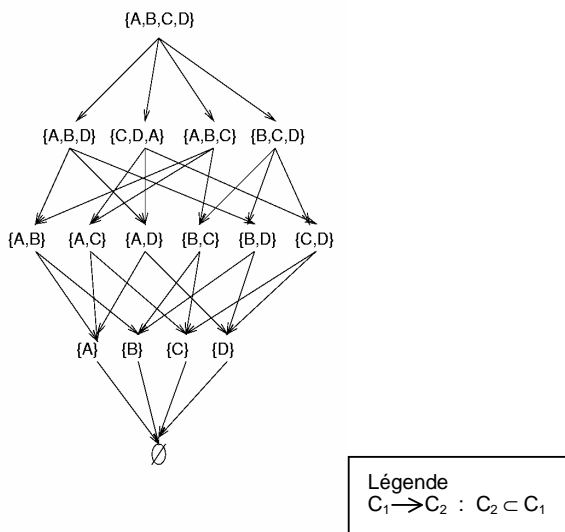


Figure 7.7 : Structure en treillis de l'espace des documents

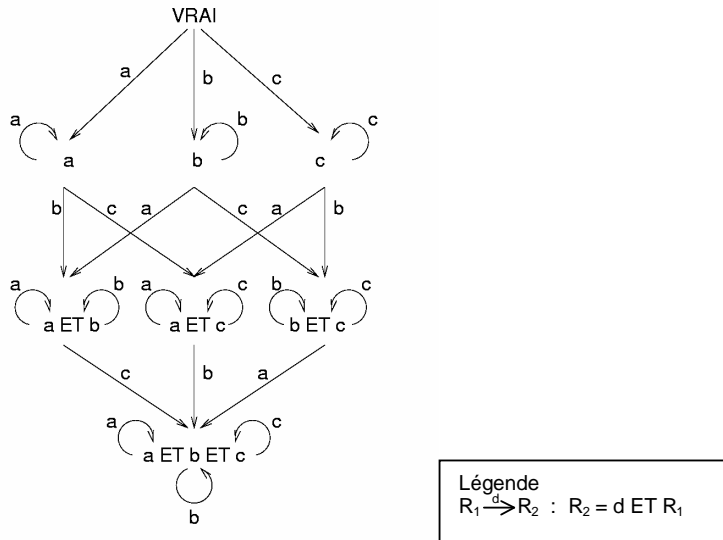


Figure 7.8 : Structure en treillis de l'espace des descripteurs

Il devient ensuite possible de déduire de la correspondance entre documents et descripteurs (cf. Tableau 7.1), la correspondance entre requêtes et corpus (cf. Tableau 7.2). De là, on remarque que certains corpus ne peuvent être obtenus par aucune requête (ex :  $\{B,C\}$ ) et que le même corpus peut être obtenu par différentes requêtes (ex : la requête  $a ET b$  et la requête  $b$ ). Ces résultats tout aussi connus qu'ils soient, ont été fort peu utilisés comme support des interactions homme-machines.

		Descripteurs		
		a	b	c
Documents	A	X		X
	B	X	X	
	C			X
	D	X		

Tableau 7.1 : Exemple de correspondance entre des documents et des descripteurs

VRAI	$\{A,B,C,D\}$
<i>a</i>	$\{A,B,D\}$
<i>b</i>	$\{B\}$
<i>c</i>	$\{A,C\}$
<i>a ET b</i>	$\{B\}$
<i>b ET c</i>	$\emptyset$
<i>c ET a</i>	$\{A\}$
<i>a ET b ET c</i>	$\emptyset$

Tableau 7.2 : Correspondance entre les requêtes et les corpus de documents (calculée à partir de l'exemple du tableau 7.1)

Claudio Carpineto et Giovanni Romano [CarpinetoEtRomano94] les ont utilisés en enlevant du graphe d'inclusion des corpus les corpus inaccessibles, obtenant ainsi un diagramme statique de généralisation/spécialisation des classes de document (cf. Figure 7.9).

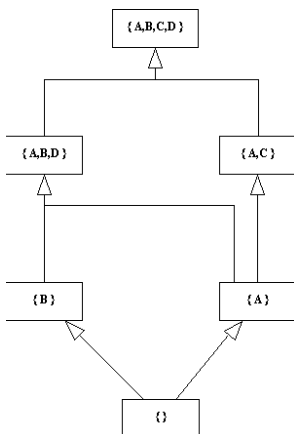


Figure 7.9 : Diagramme de classe (notation UML) dérivé de l'espace des documents

Dans notre approche, nous fusionnons dans le graphe des requêtes, celles qui décrivent le même corpus. Nous obtenons ainsi un diagramme d'état (cf. Figure 7.10) dans lequel les états correspondent à des corpus et les transitions à des requêtes élémentaires. Ces requêtes à un seul descripteur sur des corpus intermédiaires peuvent être vues comme l'ajout d'un descripteur à la requête globale : il s'agit d'une manière « d'affiner la requête » (en anglais : « query refining »).

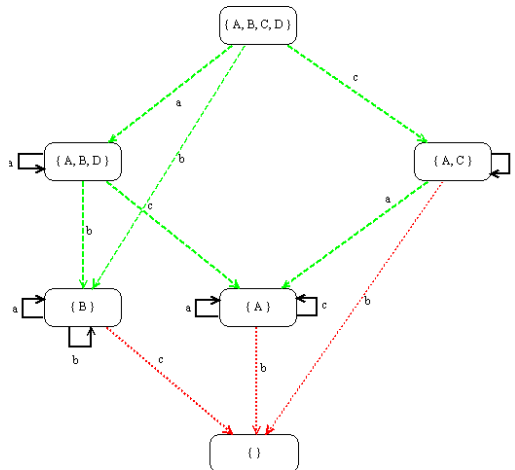


Figure 7.10 : Diagramme d'état (notation UML) dérivé de l'espace des descripteurs

Si nous reprenons le diagramme d'état précédent, dans un corpus donné chaque descripteur peut être dit :

- *impossible* : s'il mène du corpus actuel au corpus vide (ex : le descripteur  $c$  dans l'état  $\{B\}$ ),
- *connu* : s'il « boucle » sur le corpus actuel (ex : le descripteur  $a$  dans l'état  $\{B\}$ , ainsi que le descripteur  $b$  dans l'état  $\{B\}$ ),
- *possible* : dans les autres cas.

### b. Filtrer des réseaux de descripteurs

Des trois états du descripteur définis précédemment, reste à voir comment définir un filtre pour le réseau. De manière informelle notre idée est la suivante : parcourir le réseau en descendant récursivement à chaque descripteur *connu* et s'arrêter aux descripteurs *possibles* ou *impossibles*. Soyons maintenant un peu plus précis<sup>73</sup>...

**Définition :**  $_{.}getCorpus : Descriptor \rightarrow DocumentObject^n$

Fonction permettant par récursivité d'obtenir l'ensemble des objets documentaires décrits directement ou indirectement par un descripteur donné.

Note : cet algorithme est descendant car nous connaissons le descripteur et non l'objet documentaire.

$$x.getCorpus = \{y\} \leftarrow x \in Identifier \wedge x.getDOI = y$$

<sup>73</sup> Voir en annexe pour l'explication de la syntaxe utilisée dans les spécifications algébriques.

$$x.getCorpus = \{z \mid specialization(x,y) \wedge z \in y.getCorpus\} \leftarrow \\ x \in SpecializableDescriptor$$

**Définition :**  $_.getCorpus : Descriptor^n \rightarrow DocumentObject^n$

Généralisation de la fonction homonyme à une sélection de plusieurs descripteurs. L'intersection des corpus signifie qu'être décrit par une sélection de descripteurs revient à être décrit à la fois par chacun d'eux.

$$\{d_0..d_n\}.getCorpus = d_0.getCorpus \cap ..d_n.getCorpus$$

### Axiome

$$State = \{KNOWN, POSSIBLE, IMPOSSIBLE\}$$

**Définition :**  $_.getState(\_) : Descriptor \times DocumentObject^n \rightarrow State$

Fonction attribuant un état (*connu*, *possible*, *impossible*) à un descripteur pour un corpus  $C$  donné. Plus précisément, on dira que le descripteur est *connu* si le corpus du descripteur contient  $C$  et que  $C$  est non vide. Il sera *impossible* si le corpus du descripteur a une intersection vide avec  $C$ . Il sera *possible* dans les autres cas.

$$x.getState(C) = KNOWN \leftarrow C \neq \emptyset \wedge C \subseteq x.getCorpus$$

$$x.getState(C) = IMPOSSIBLE \leftarrow C \cap x.getCorpus = \emptyset$$

$$x.getState(C) = POSSIBLE \leftarrow$$

$$x.getState(C) \neq IMPOSSIBLE \wedge x.getState(C) \neq KNOWN$$

OPTIMISATION : On ne calculera *getCorpus* qu'une fois et on procédera par élimination en parcourant  $C$  de conserve avec *getCorpus* pour y trouver un contre-

exemple à *KNOWN* (un élément de *C* qui n'est pas dans *getCorpus*) et à *IMPOSSIBLE* (un élément de *C* qui est dans *getCorpus*).

**Définition :**  $\_getFilter(\_) : Descriptor \times DocumentObject^n \rightarrow (Descriptor \times State)^n$

Fonction de filtrage descendant récursivement dans les descripteurs connus et s'arrêtant aux descripteurs possibles et impossibles.

$$x.getFilter(C) = \{(x, IMPOSSIBLE)\} \leftarrow x.getState(C) = IMPOSSIBLE$$

$$x.getFilter(C) = \{(x, POSSIBLE)\} \leftarrow x.getState(C) = POSSIBLE$$

$$x.getFilter(C) = \{(x, KNOWN)\} \cup \{(z,s) \mid specialization(x,y) \wedge (z,s) \in y.getFilter(C)\} \\ \leftarrow x.getState(C) = KNOWN$$

**Définition :**  $\_getDescriptionContext(\_) : Facet^n \times Descriptor^{n \times n} \rightarrow (Descriptor \times State)^{n \times n}$

Obtention du contexte de description pour un ensemble de facettes et de sélections associées. On calcule d'abord le corpus global par intersection des corpus de chaque facette. Ensuite, on applique à chaque facette le filtre correspondant au corpus global.

$$\{f_0, \dots, f_n\}.getDescriptionContext(\{S_0, \dots, S_n\}) = \{f_0.getFilter(C), \dots, f_n.getFilter(C)\} \\ \leftarrow C = f_0.getCorpus(S_0) \cap \dots \cap f_n.getCorpus(S_n)$$

Optimisation 1 : L'utilisateur ne pouvant modifier qu'une sélection à la fois, il n'est pas nécessaire de recalculer les corpus des autres facettes. On conservera donc un « cache » pour chaque facette. Celui-ci ne sera remis à jour qu'à la suite d'une modification de sélection.

Optimisation 2 : Chaque facette pouvant se situer sur un serveur différent, les appels aux méthodes *getFilter* se feront de manière parallèle.

Les figures 7.11 et 7.12 illustrent l'obtention du contexte de description et ses optimisations.

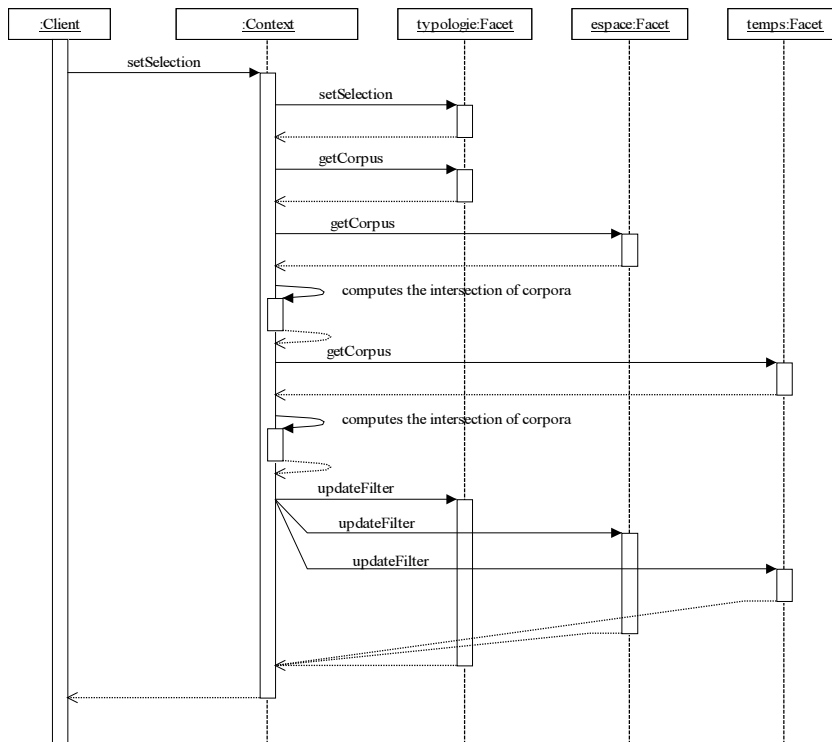


Figure 7.11 : Mise à jour des filtres dans les facettes « typologie », « espace » et « temps » après changement de sélection dans la facette typologie (diagramme de séquence UML).

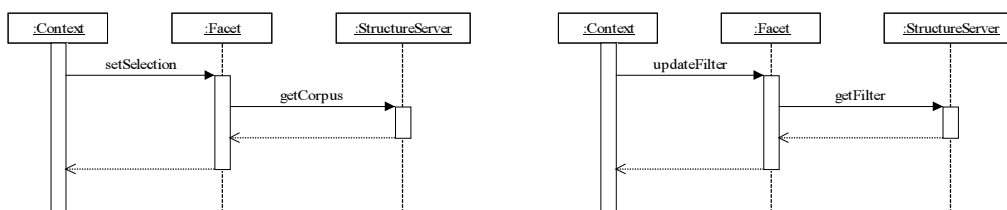


Figure 7.12 : Requêtes envoyées au serveur lors de la mise à jour des filtres (cf. Figure précédente).



**c. Scénario de recherche de documents**

Dans cette section, nous suivrons pas à pas un scénario d'interactions homme-machine mobilisées lors d'une recherche de document. Les schémas présenteront le réseau de description de la Figure 7.13 tel que filtré par le système *Porphyre* au cours d'une navigation.

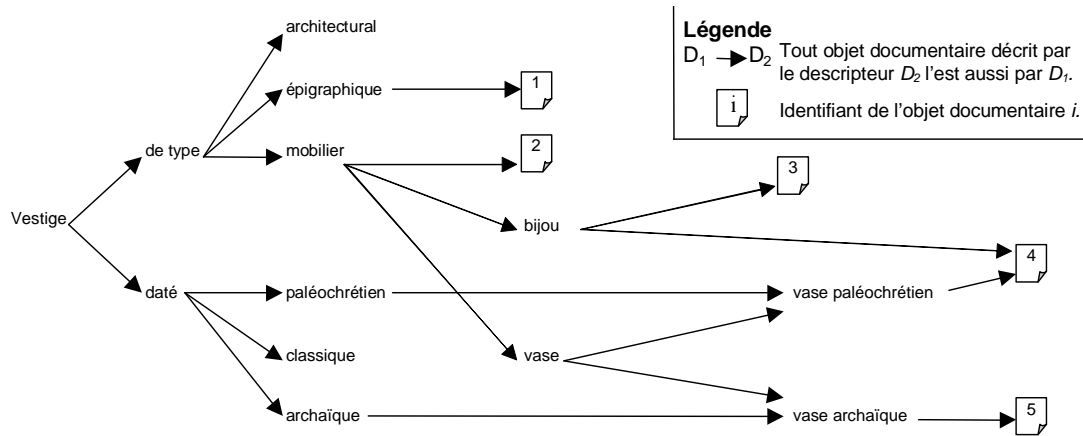
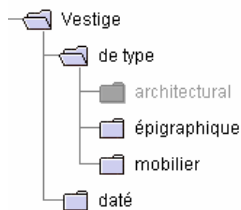
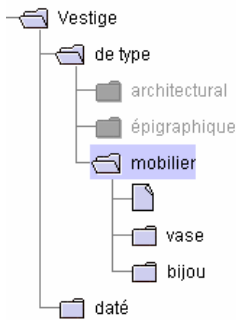


Figure 7.13 : Exemple de réseau de description

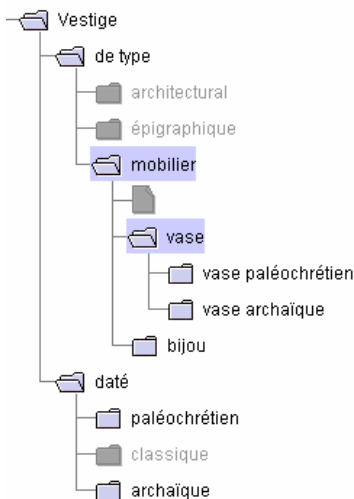
**Étape 1 :** Le corpus global traite de « vestige typé ». Les corpus plus spécialisés traitent de « vestige daté », de « vestige de type épigraphique » ou de « vestige de type mobilier » mais pas de « vestige de type architectural » (cette description ne correspond en effet à aucun document du corpus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



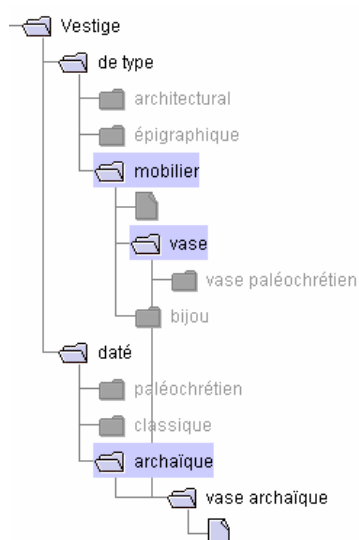
**Étape 2 :** Le corpus sélectionné traite de « vestige de type mobilier ». Cette description correspond exactement à celle de l'un des documents. Celui-ci est affiché. Des corpus plus spécialisés traitent de « bijou » ou de « vase ». Aucun ne traite de « vestige de type architectural » ou « épigraphique ». Après sélection par l'utilisateur de « vase », le système passe à l'étape 3.



**Étape 3 :** Le corpus sélectionné traite de « vestige de type vase (mobilier) » et de « vestige daté ». On remarque que le fait qu'il soit « daté » est inféré par l'ordinateur (car tous les documents du corpus traitant de « vase » traitent de « vestige daté »). Les corpus plus spécialisés traitent de « vase paléochrétien », de « vase archaïque », de « paléochrétien » ou « d'archaïque ». Aucun ne traite de vestiges de type « architectural », « épigraphique », « classique ». Le document qui était possible à l'étape précédente ne l'est plus. Après sélection par l'utilisateur « d'archaïque », le système passe à l'étape 4.



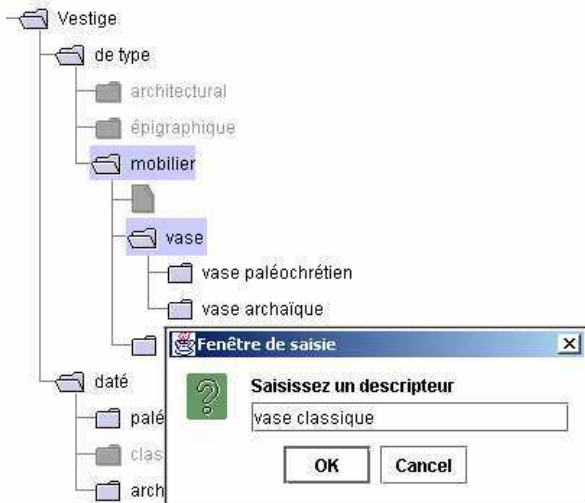
**Étape 4 :** Le corpus sélectionné traite de « vestige de type vase (mobilier) daté de l'époque archaïque ». Ce corpus ne contient qu'un seul document. Celui-ci est affiché.



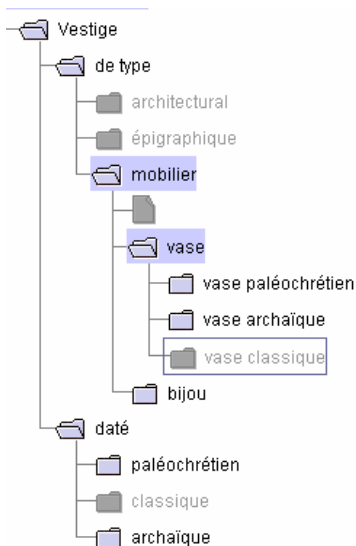
#### **d. Scénario d'indexation de documents**

Dans ce scénario, nous allons indexer un nouveau document traitant de « vase classique ». Tout d'abord, l'utilisateur doit trouver pour le nouveau document un corpus qui lui corresponde (mais pas ses sous-corpus). Comme le document traite de « mobilier » et de « vase », l'utilisateur peut suivre les étapes 1 et 2 (de la section précédente). Nous nous retrouvons alors à l'étape 3'.

**Étape 3'** : Aucun des sous-corpus proposés (« vase paléochrétien », « vase archaïque », « paléochrétien » et « archaïque ») ne correspond au document à indexer. L'utilisateur crée donc un nouveau descripteur « vase classique » comme spécialisation de « vase », le système passe alors à l'étape 4'.

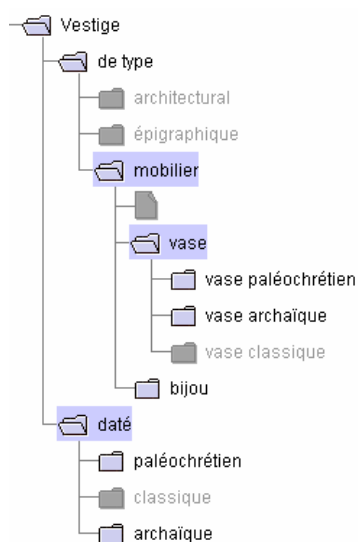


**Étape 4'** : Le nouveau descripteur « vase classique » apparaît comme n'étant affecté à aucun document. C'est alors que l'utilisateur crée l'identificateur de document « 6 » comme spécialisation de « vase classique ». Le système passe à l'étape 5'.

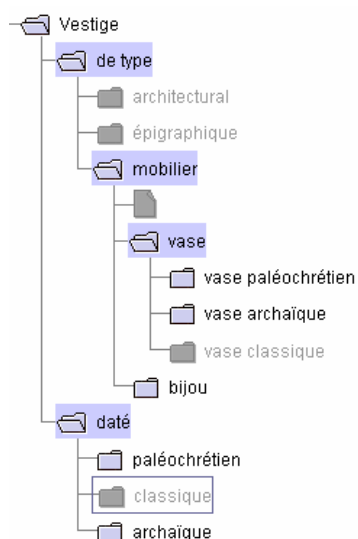


## CHAPITRE 7. RÉSEAUX DE DESCRIPTION\*

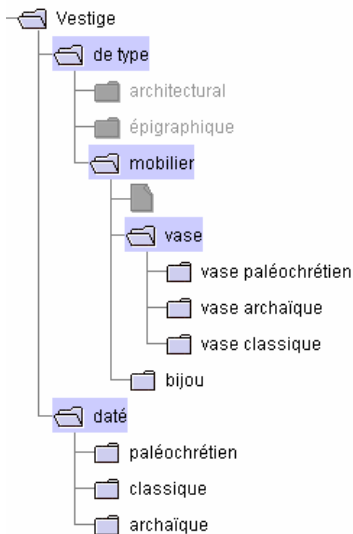
**Étape 5'** : Par conséquent, le système montre que « vase classique » décrit un corpus et que « daté » ne décrit plus l'ensemble des documents décrits par « vase ». Afin de comprendre cette incohérence potentielle, l'utilisateur sélectionne « daté ». Le système passe à l'étape 6'.



**Étape 6'** : Il apparaît que « vase classique » n'est pas déclaré comme « daté » et que « classique » n'est affecté à aucun document. L'utilisateur met alors le modèle à jour en reliant « vase classique » à « classique ». Le système passe à l'étape 7'.



**Étape 7'** : L'utilisateur a terminé d'indexer le nouveau document et « tout est pour le mieux dans le meilleur des mondes possibles »...



#### 4. Bilan

Nous avons tout d'abord rappelé que l'organisation des bibliothèques (comme celle des documents) est en général basée sur un modèle arborescent, probablement en raison de l'influence de la philosophie d'Aristote sur notre manière de voir la connaissance. De manière à sortir de ce schéma trop simpliste, nous avons étudié certaines alternatives. Cependant, du fait qu'elles s'appuient toujours sur un modèle du domaine considéré comme fixe et extérieur, ces alternatives nous ont semblé telles quelles inapplicables à notre approche : celle d'une modélisation dynamique effectuée par les experts eux-mêmes et non par des tiers.

Nous avons ensuite proposé un modèle appelé « réseau de description » permettant à chaque expert de superposer au corpus sa propre structure, son propre point de vue. Une définition a alors été donnée de manière d'abord informelle puis à l'aide de spécifications algébriques.

Enfin nous avons vu, de manière formelle, puis sur des scénarii, comment on pouvait instrumenter l'expert dans sa tâche d'arpentage de l'espace documentaire grâce à un mécanisme de filtre.