Du Content management system au Concept management system Une lecture critique de l'article de C. Lagoze et al. " Qu'est-ce qu'une bibliothèque numérique au juste ? "

Jean-Daniel Zeller

▶ **To cite this version:**

Jean-Daniel Zeller. Du Content management system au Concept management system Une lecture critique de l'article de C. Lagoze et al. " Qu'est-ce qu'une bibliothèque numérique au juste ? ". 2007. sic_00154352

HAL Id: sic_00154352

https://archivesic.ccsd.cnrs.fr/sic_00154352

Preprint submitted on 13 Jun 2007

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.

Du Content management system au Concept management system

(De le gestion de contenu à la gestion des concepts ?)

Une lecture critique de l'article de C. Lagoze et al.
« Qu'est-ce qu'une bibliothèque numérique au juste ? »[1]

Résumé

A partir d'une problématique d'enseignement (en vue d'un télé-enseignement), l'auteur analyse et propose de compléter le modèle de base de contenu proposé par Carl Lagoze et al. dans le logiciel FEDORA. Les nouvelles structures proposées sont : le fait élémentaire - la définition - le fragment - l'annotation - le concept. L'analyse soulève également la question de la granularité de la connaissance et de la charge cognitive qui lui est associée.

Abstract

Starting from of teaching problematic (in an e-learning scope), the author analyzes and proposes to complete the basic content management system proposed by Carl Lagoze and al. in the software FEDORA. The new suggested structures are: the elementary fact - the definition - the fragment - the annotation - the concept. The analysis also raises the question of the granularity of knowledge and the cognitive load which is associated.

Mots-clés

Sciences de l'information, Bibliothèque électronique, Entrepôt donnée, Gestion de contenu, Système information, Modèle de donnée, Métadonnée, Interopérabilité, Logiciel libre, Fedora, Charge cognitive, Gestion des connaissances

Keywords

Information Science, Digital library, Data warehouse, Content management, Information system, Data model, Metadata, Interoperability, Open source software, Fedora, Cognitive load , Knowledge management.

Introduction

L'article de Carl Lagoze propose un système qui semble pouvoir résoudre une question devenue de plus en plus aiguë dans le cadre de l'enseignement. J'exprime cette question ci-dessous de manière personnelle, en rapport avec mon activité d'enseignement depuis plus de dix ans dans un domaine de connaissances particulièrement évolutif, puisqu'il s'agit de l'archivage des documents électroniques.

En 1995, la bibliographie spécifique sur le sujet se composait de cinq à six ouvrages, largement redondants, et d'une cinquantaines d'articles, la plupart anglo-saxons. Aujourd'hui, je suis dans l'incapacité de suivre toutes les publications dans le domaine, publications par

[1] Accessible sur le site de l'ARTIST : http://dilib.inist.fr/~ducloy/AMETIST/public/0/LAGOZE/

ailleurs largement répertoriées dans plus d'une cinquantaine de sites Internet (dont certains ont disparus depuis). Une interrogation sur Google [le 2.1.2007] donne comme résultat :

Mot clé	**Réf. web**	**Réf. francophones**
« Digital document »	965'000	715
« Digital archiving »	380'000	697
« Document numérique »	178'000	167'000
« Archives électroniques »	44'700	41'900
« Archives numériques »	36'800	35'500
TOTAUX	1'604'500	245'812

Ce volume rend de plus en plus difficile l'inclusion des nouvelles percées théoriques et/ou pratiques de ce champs de connaissance dans un module d'enseignement, forcément limité dans le temps disponible (en l'occurrence 32 heures de cours). Il apparaît assez légitime de généraliser cette problématique à tous les domaines de connaissance en évolution rapide.

Dans ce contexte, j'étais à la recherche d'un outils informatique à la fois suffisamment simple pour permettre une appropriation rapide à un non-informaticien, et suffisamment sophistiqué pour permettre une mise en relation conceptuelle de différents objets numériques qui fasse sens. Une autre contrainte était pour moi la possibilité d'utiliser une base de connaissances pour des populations d'étudiants différentes, sans avoir à la reformuler (reformater) dans un support de cours qui n'est jamais à jour ni jamais totalement adéquat, mais qui constitue plutôt un réservoir d'objets qui me permette de naviguer à différents niveaux dans la base selon différentes modalités, dont je cite trois exemples ci-dessous.

Un parcours introductif (flèche A)

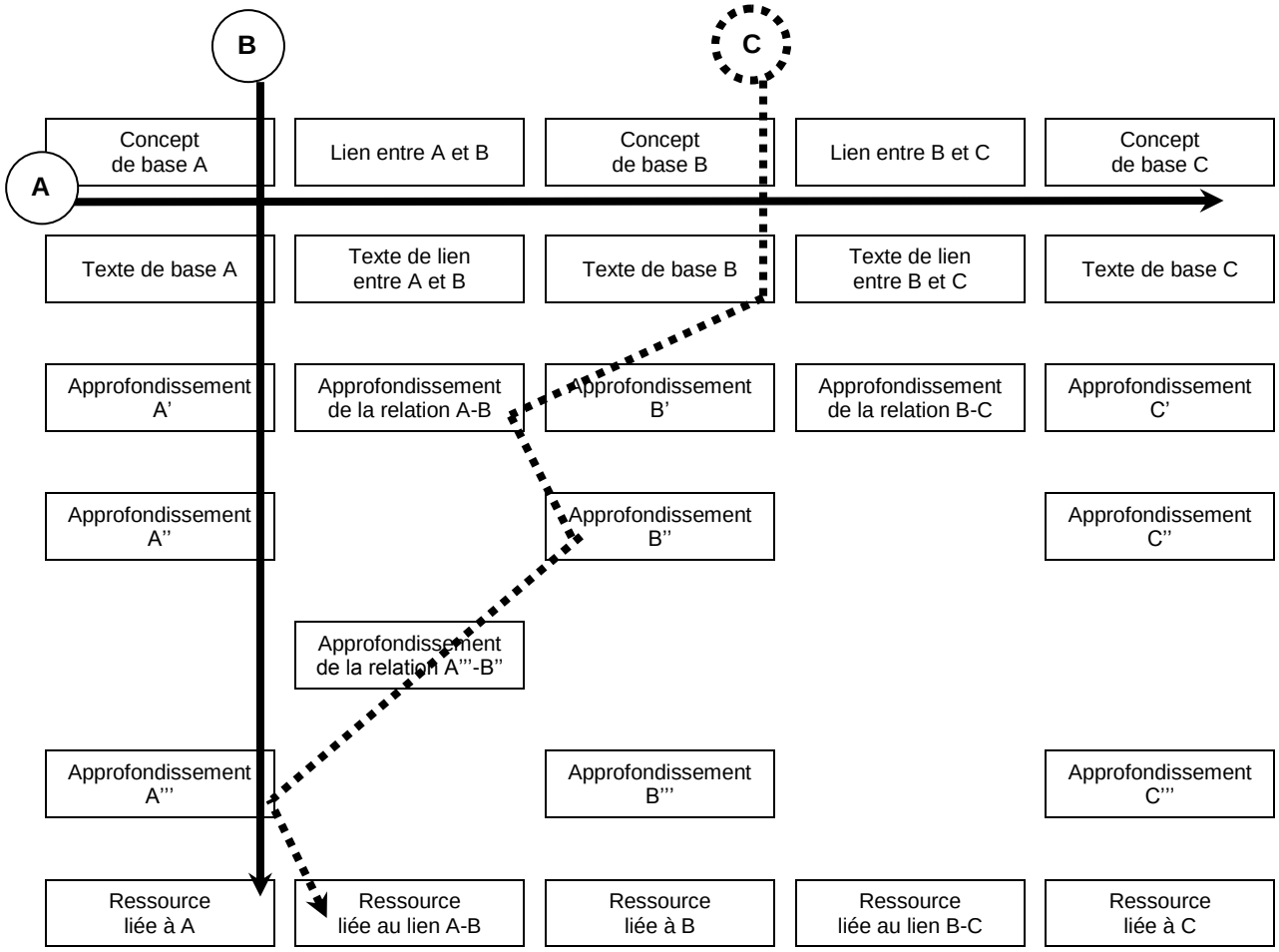
Soit une présentation systématique (et/ou chronologique) des concepts clés.

a) Un approfondissement de certains concepts (flèche B)

Concepts pouvant être plus ou moins approfondis selon les besoins des publics concernés.

b) Un parcours « à la carte » (flèche C)

Au gré des besoins non plus d'un cursus (savoirs obligatoires) mais d'un projet ou d'un atelier (savoirs en situation).



Mon exploration dans ce domaine m'a amené à examiner les ***logiciels de cartes heuristiques***. Cependant la plupart de ceux-ci permettent de gérer des liens graphiques mais pas de gérer un système conceptuel complexe, ni de relier de manière systématique les concepts représentés aux objets numériques qui les fondent (si ce n'est par un hyperlien « bête »).

L'autre versant de ma recherche se porta du côté des ***logiciels de télé-enseignement***. Ici, l'appareillage conceptuel était en général consistant mais systématiquement organisés en objets pédagogiques (cas, cours, exercice, lecture, etc.) et pas en objets de connaissance. Ce biais pédagogique s'exprime également par l'attachement d'un objet pédagogique à des niveaux scolaires pré-déterminés[2]. D'autre part, même si une tendance à la normalisation se

[2] Même les projets novateurs comme la Wikiversité sont restés attachés à l'organisation en facultés et départements [http://fr.wikiversity.org/wiki/Wikiversit%C3%A9:Organisation_des_enseignements], tout en

fait sentir (norme SCORM[3] par exemple), elle s'attache plus à l'interopérabilité informatique qu'à l'interopérabilité conceptuelle. Ceci allait à l'encontre de mon désir de base puisque je cherchais justement à court-circuiter ces catégories de type niveau de cours en créant une base de connaissance destinée à un usage multiforme (en terme universitaire, aussi bien à des fins d'enseignement qu'à des fins de recherche). Enfin, un dernier défaut de ces systèmes est la typologie des objets utilisés, qui entrelacent souvent des types liés aux formats informatiques à des types logiques et/ou conceptuels.

J'en était là de mes recherches, qui m'amenaient au désespoir, quand l'article de Lagoze et al. parvint à ma connaissance. Le principe de base qui le sous-tend me semble une idée pragmatique extrêmement féconde qui consiste à séparer dans la base les objets numériques, leurs métadonnées et leur liens réciproques, tout en maintenant un unique langage informatique (XML) qui les rend éminemment interopérables via un interface web mondialement disponible. Le fait de dissocier les objets et leur caractéristiques de formats informatiques (MIME) qui sont facilement traitables par les logiciels, de leurs descriptions (métadonnées et annotations) qui sont elles aisément manipulables humainement au niveau du concept, apparaît comme la base de la performance du système.

Du contenu au contexte, et du contexte au concept

En examinant de plus près les fonctionnalités de FEDORA, logiciel open source que l'équipe de Lagoze présente comme un *Content Management System* (CMS) on s'aperçoit que celui-ci, de par les liens « qualifiés » qu'il permet d'établir entre les objets numériques présents dans la base, peut en fait également être appelé un *Context Management System*. Mon hypothèse est que, moyennant certains enrichissements que je détaille plus loin, il pourrait même se transformer en un *Concept Management System* et constituer un modèle de base de connaissances extrêmement puissant et utile pour l'enseignement de l'avenir.

La question de la volumétrie conceptuelle

Indépendamment de la création de nouveaux types d'objets que je propose plus loin, le modèle présente un inconvénient sur le plan pratique. S'il est tout à fait légitime au plan informatique de gérer avec un même modèle de métadonnées une image (ou un jeu d'images) un article scientifique, ou une monographie ; cette équivalence n'est plus praticable sur le plan conceptuel car ces différents éléments représentent une charge cognitive différente. Ceci est illustré par les deux figures ci-dessous. La première reproduit une représentation à l'image de celle utilisée dans l'article original, la seconde représente les mêmes éléments mais pondérés par le « poids cognitif » des différents objets. Comme il n'existe pas d'estimation de ce poids j'ai par convention utilisé leur équivalent en volume (en Ko). Bien sur, ce volume n'est pas l'exacte représentation de la charge cognitive induite par les objets mais il donne une image de la disproportion existante, qui est largement suffisante pour la réflexion en l'état.

identifiant le problème que cette organisation pose par rapport aux savoirs transversaux. Voir la discussion sur la page colloque [http://fr.wikiuniversity.org/wiki/WikiversitÃ© :Colloque]

[3] Voir le site de SCORM : [http://fr.wikipedia.org/wiki/SCORM] & [http://www.adlnet.gov/scorm/index.cfm]

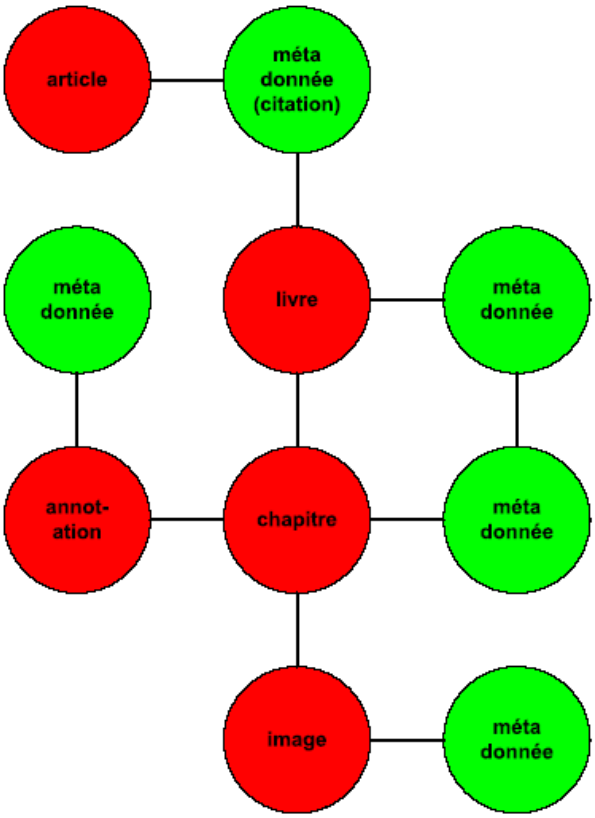


Fig. 1 : Relations conceptuelles entre les différents objets de la base de connaissance (selon C. Lagoze)

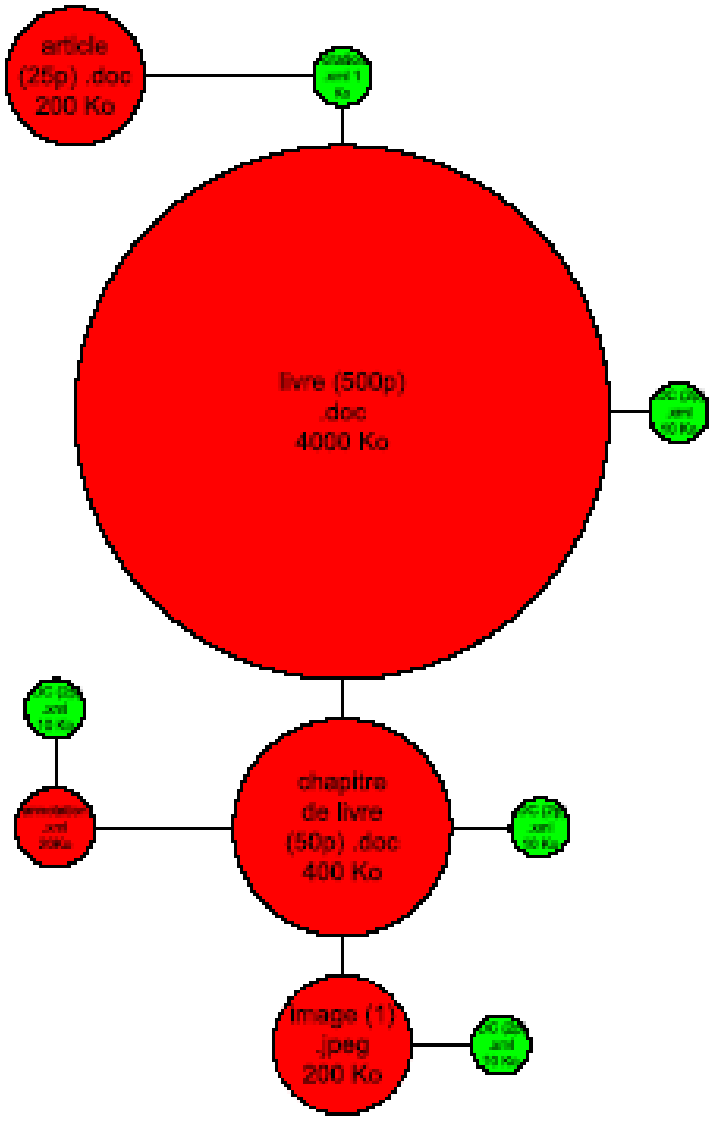


Fig. 2 : Relations conceptuelles « pondérée » entre les différents objets de la base de connaissances (compte tenu de la disproportion effective, l'échelle n'est pas exactement respectée)

La comparaison des deux figures montre à l'évidence que la monographie présente une charge relative disproportionnée par rapport aux autres objets. Il est donc nécessaire de le « métaboliser » pour le rendre plus « digeste », par exemple en le divisant en chapitre. Cette opération met cependant en cause la cohérence du discours et pose la problématique de la bonne granularité de la connaissance[4]. Formulée idéalement, c'est répondre à la question « quel est la quantité d'information optimale qui représente une notion auto-suffisante ? ». Je laisserais cette question ouverte ici, son approfondissement devant faire l'objet d'une réflexion pour elle-même.

[4] Sur cette question, on consultera avec profit les considérations ergonomiques de Michel Cartier dans sa publication *Le Web de 3ème génération* [http://www.michelcartier.com/McArticleB.php3?id_article=271]

De nouveaux types d'objet pour l'exhaustivité de la base de connaissance

Les nouveaux objets décrits ci-dessous peuvent parfaitement rentrer dans le modèle XML décrit dans l'article cité pour décrire l'annotation, ils sont différenciés en fonction de leur utilité conceptuelle et pourraient faire simplement l'objet d'un jeu de métadonnées descriptives légèrement différentes.

Les différents types proposés sont les suivants :

1/ le fait élémentaire
2/ la définition
3/ le fragment
4/ l'annotation (déjà décrit par Lagoze)
5/ le concept

Je définis ci-dessous les propriété de ces types puis j'explique in fine quels sont les liens qui les lient et en font l'utilité dans la construction d'une base de connaissances.

Le fait élémentaire

Un fait élémentaire est une donnée ou un jeu de données mesuré et/ou calculable explicitement, dont la véracité est communément admise (cette dernière approximation est liée à son utilisation en sciences humaines où la validité absolue exigée en science dure n'est pas forcément atteignable). Comme même les vérités scientifiques changent, un fait élémentaire doit être daté. Cette datation permet également d'établir les variations chronologique d'un même type de fait. Ainsi, le fait « 60% des utilisateurs de Google n'utilisent qu'un seul mot pour effectuer une recherche » (en 2005) pourrait être complété par un fait ultérieur comme « 80% des étudiants français ayant suivi un cursus en Info-doc utilisent une recherche booléenne dans les instruments de recherche » (en 2010). Un fait élémentaire peut soit être extrait d'un objet déjà existant dans le système, auquel cas ses métadonnées renverront à cette source (un lien selon Lagoze), soit il peut être issu d'une recherche en cours, auquel cas il peut être déposé dans le système comme un objet en soi, qui sera utilisé pour un usage ultérieur (une publication par exemple).

La définition

Si le sens de la définition est clair et commun (Proposition composée de termes censés connus et permettant de reconnaître un objet par ses principaux caractères, Grand dictionnaire terminologique), dans le cadre d'un tel système, il s'agit de définition « qualifiée ». Ainsi pour un mot comme « inconscient » le système pourrait contenir par exemple trois définitions différentes, renvoyant aux définitions de ce terme données par Freud, Jung ou Adler, et liées bien entendu aux publications de ces auteurs où elles figurent. Le fait d'inclure ces définitions en tant qu'objet dans le système plutôt que de créer des liens vers des dictionnaires externes (comme Wikipedia par exemple) est justifié par le fait que des acceptions nouvelles ou des néologismes peuvent être créées en relation avec de nouveaux objets de la base. L'ajout de liens entre ces définitions et les autres objets de la base permet la génération « spontanée » d'ontologies. Ainsi, une définition peut renvoyer au texte où elle figure, à une autre définition, à un concept (en tant que « nœud ontologique », voir plus loin), à un fait élémentaire, à un

auteur, etc. Ici également, un datation est nécessaire puisque des termes définis une fois peuvent voir leur sens évoluer dans le temps ou selon les auteurs.

Ce travail de « metadonnisation », consistant en l'ajout systématique de métadonnées à un type d'objet numérique, devrait permettre d'établir des terminologies comparatives des plus instructives. Dans ce sens, l'indexation matière des documentalistes est un cas particulier de métadonnisation.

Le fragment

Un fragment est n'importe quelle partie d'un objet numérique préexistant faisant sens par lui-même (autoconsistance). Mais un fragment peut aussi être une assertion existant hors d'un autre objet numérique (hypothèse de travail, intuition, groupement de faits élémentaires). Il a donc forcément un auteur et une date. Je propose le terme fragment pour permettre un usage le plus générique possible. Dans mon esprit le fragment correspond à une molécule d'information alors que la donnée (définie comme un triplet : entité, attribut, valeur)[5] en serait l'atome. Je trouve cependant dangereux de pousser l'analogie chimique plus loin, ce qui pourrait amener à des conclusions fallacieuses. De même, on pourrait utiliser le terme de « grain » d'information….(réf), ce qui serait peut-être plus satisfaisant, la granularité pouvant être variable selon l'usage visé et selon la « métabolisation » du savoir envisagée. Cependant cela reste pour l'instant trop contraignant et je préfère garder le terme plus générique de fragment, quitte à ce qu'avec l'expérience d'une base d'objets numériques réelle, elle soit affinée/développée.

Dans le cadre d'un fragment d'un objet préexistant, il peut être utile que, en plus de l'auteur initial, on détermine un auteur de la fragmentation avec les métadonnées indiquant le « sens » de la fragmentation, par exemple « ce fragment fait sens dans le cadre d'un réflexion sur le document numérique », alors que dans un autre cadre, l'indexation par exemple, c'est un autre fragment du même texte qui peut faire sens.

Un fragment peut également constituer un lien contextuel explicite fait entre plusieurs objets numériques de la base, quoique l'on se trouve dans ce cas plutôt en face d'une annotation telle que la défini Lagoze. La mise en relation de plusieurs fragments de ce type avec un concept, pourrait constituer de manière quasi automatique la bibliographie commentée de ce concept (voir plus loin le terme concept).

L'annotation

Lagoze défini l'annotation comme n'importe quel type de commentaire rattaché à un autre objets numériques, de manière très générique. Ce qui est une définition acceptable et non contradictoire avec les autres objets plus spécifiques que je propose. Son intérêt réside dans la possibilité potentielle qu'elle peut apporter quand elle est reliée avec les autres objets que je suis en train de définir. Lié à des concepts et des définitions l'annotation prend place dans une constellation qui permet de relier un concept à des citations, formant une ébauche d'un champ conceptuel, comme l'illustre le schéma ci-dessous.

[5] cf. [Redman, 1998], plus particulièrement le chapitre 12 : « Qu'est-ce qu'une donnée ? »

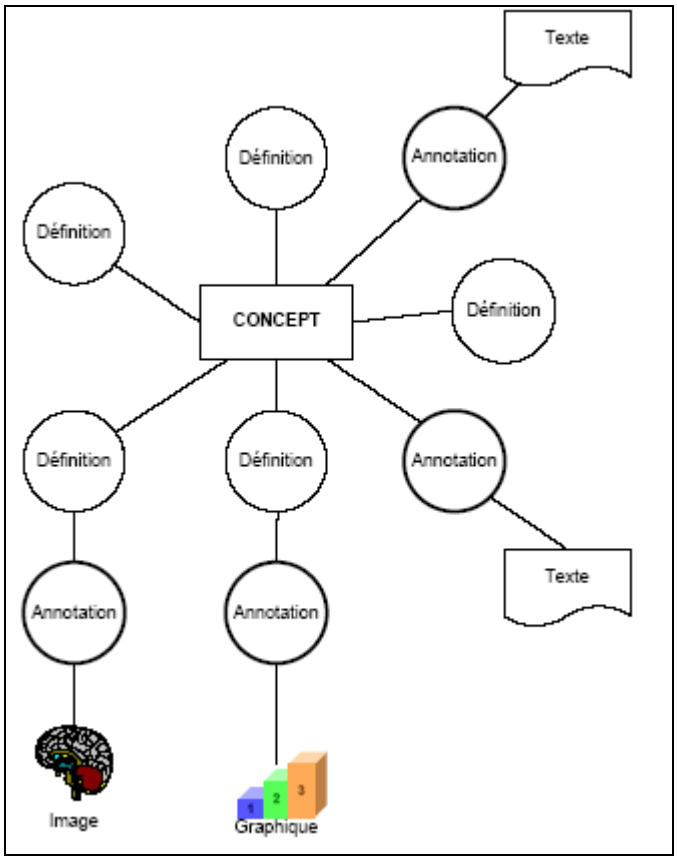


Fig. 3 : Réseau d'annotations

Le concept

Selon le dictionnaire, un concept est une idée ou représentation de l'esprit qui abrège et résume une multiplicité d'objets empiriques ou mentaux par abstraction et généralisation de traits communs identifiables. Dans le cadre de ma proposition un concept est le développement encyclopédique d'une ou de plusieurs définitions. Il représente la coordination de différents éléments de la base constituant un champs conceptuel est sera donc fréquemment représenté graphiquement comme un champ de relations (ce qui pose le problème de son adéquation avec le langage XML préconisé comme standard de représentation des objets dans le système de Lagoze).

Un concept communément établi sur le plan paradigmatique, comme par exemple le théorème de Pythagore ou le principe d'incertitude d'Heisenberg, sera plus fortement corrélé avec des définitions et des textes canoniques, comme représenté par le schéma ci-dessous :

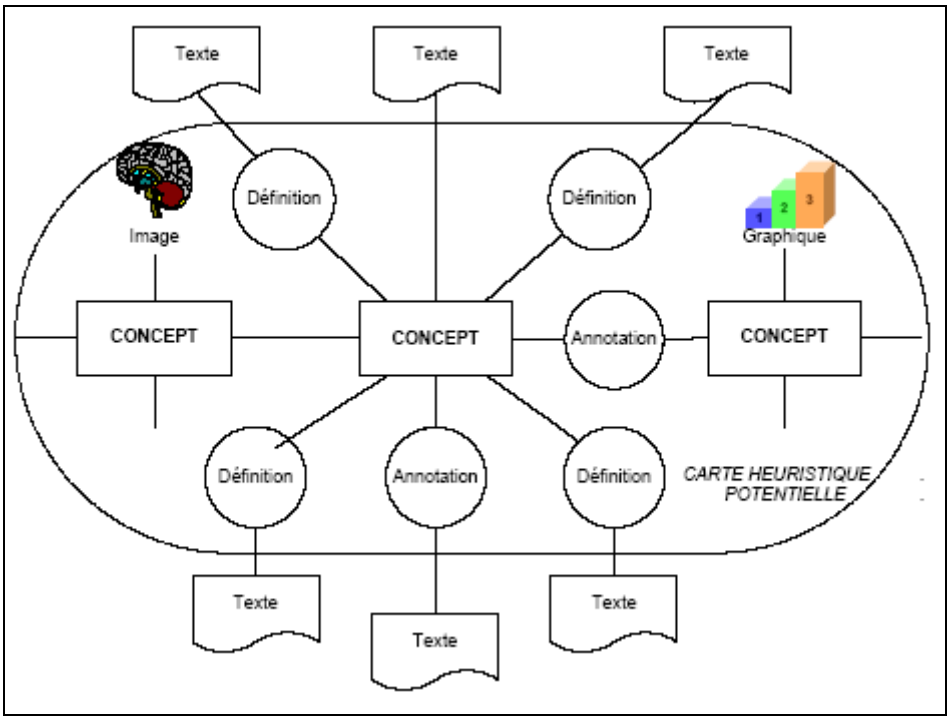


Fig. 4 : Réseau conceptuel

Un concept en formation (hypothèse de travail, nouveau paradigme, etc.) sera lui plutôt relié à des fragments, des annotations ou des définitions de type néologisme, mais selon le même type de constellation.

Dans les deux cas, le concept représente en quelque sorte un « noyau d'ontologie », la concaténation des concepts reliés entre eux de près ou de loin aboutissant théoriquement à l'ontologie complète d'un champs conceptuel donné. Dans une telle base, il serait intéressant de comparer l'ontologie issue d'une une analyse des corrélations effectuées manuellement au fil de sa constitution avec les ontologies extraites au moyen d'algorithmes sur le contenu des textes de la même base. A mes yeux, l'avantage de l'ontologie « issue des liens » est sa capacité d'évolution spontanée au fil de l'ajout d'autres textes et d'autres liens dans la base tandis que la méthode algorithmique oblige la constitution d'un corpus qui induit assez régulièrement des biais (rechercher références) et sont difficile à mettre à jour[6].

[6] **V**oir sur cette problématique : Claude Chrisment, Françoise Genova, Nathalie Hernandez et Josiane Mothe ***D'un thesaurus vers une ontologie de domaine pour l'exploration d'un corpus,*** Ametist, no 0, 2006. http://ametist.demo.inist.fr/document.php?id=152

Conclusion(s)

Les propositions ci-dessus sont des hypothèses de travail. Il reste à les valider par :

a) La création de modèles XML correspondant aux objets proposés.
b) La création d'une base de connaissance d'un domaine délimité avec ces mêmes types d'objets (les archives numériques par exemple).
c) L'exploitation d'une telle base dans le cadre d'un (et si possible plusieurs) cursus d'enseignement.

Jean-Daniel ZELLER
Enseignant à la
Haute école de gestion
Dépt. Information documentaire
Genève

Bibliographie

[Cartier, 2003]
Michel Cartier, *Le Web de 3ème génération*, 2003
[http://www.michelcartier.com/McArticleB.php3?id_article=271]

[Chrisment, 2006]]
Claude Chrisment, Françoise Genova, Nathalie Hernandez et Josiane Mothe, *D'un thesaurus vers une ontologie de domaine pour l'exploration d'un corpus*, Ametist, no 0, 2006.
http://ametist.demo.inist.fr/document.php?id=152

[Lagoze, 2006]
Carl Lagoze et al., *Qu'est-ce qu'une bibliothèque numérique, au juste?*
"*Au-delà des fonctions recherche et accès dans la National Science Digital Library*",
AMETIST, no 0, sept. 2006
http://dilib.inist.fr/~ducloy/AMETIST/public/0/LAGOZE/

[Paquette, 2002]
Paquette, G. La modélisation des connaissances et des compétences, pour concevoir et apprendre, Presses de l'Université du Québec, 2002

[Redman, 1998]
Redman T., *La qualité des données à l'âge de l'information*, Paris, InterEditions, 1998.