Rapport Final Arkeotek

Valorisation du projet Arkeotek : atelier international sur « Indexation et connaissances en sciences humaines »

Juillet 2005 – Juillet 2006

Coordinatrice : N.Aussenac-Gilles (IRIT, UMR 5505)

Equipes partenaires : V. Roux (Préhistoire et Technologie, UMR 7055),
P. Blasco (Editions Epistèmes)

Projet également mené avec les équipes du projet Porphyry :
A. Iacovella (EPML STIC 50)
S. Calabretto (LIRIS UMR 5205),
A. Bénel (UTT)

Cognitique / Programme Société de l'information du CNRS
Appel 2005 « Valorisation »

Rapport Interne IRIT/RR--2006-02--FR

Résumé

Avec l'irruption de l'édition électronique, l'archivage des données scientifiques peut enfin être pensé dans la continuité du travail d'édition des constructions théoriques. La question de l'indexation des données et les paradigmes de consultation sont dès lors profondément modifiés. Pour tirer parti de cette évolution conceptuelle, l'apport des techniques d'ingénierie de la connaissance peut être déterminant. Le projet Arkeotek offre un cadre pour examiner l'utilisation de tels outils d'ingénierie des connaissances à des fins de création de vues dynamiques, synthétiques et hiérarchisées, de documents et de données scientifiques regroupés au sein d'une base de connaissances. Une ontologie des domaines scientifiques concernés est utilisée pour l'indexation des documents. Compte tenu du format original de représentation des documents, la recherche d'information et l'édition des résultats peut alors s'appuyer à la fois sur le contenu, la sémantique des documents et leur structure, elle-même révélatrice de l'argumentation scientifique.

Ce type d'approche se développe dans différents contextes en sciences humaines, que ce soit pour favoriser l'annotation de documents pour leurs auteurs ou lecteurs, l'exploitation de grandes bases documentaires, la navigation ou la recherche d'information au sein de collections. Ainsi, le projet Porphyry fait également appel à des modèles, proches d'ontologies, pour aider un chercheur en archéologie à annoter ses propres documents. A cet état d'avancement de la réflexion, il nous a semblé intéressant de mettre en place un atelier au cours duquel seraient confrontées différentes approches ayant en commun de favoriser le travail du chercheur en sciences humaines. Nous avons souhaité que les équipes concernées développent des solutions techniques dans un cadre véritablement interdisciplinaire, les chercheurs en sciences humaines menant une véritable réflexion sur la structuration de leur savoir, sa diffusion et les modes d'accès.

Ce rapport présente donc les éléments préparatoires à un atelier international que nous organisons, en collaboration avec le projet Porphyry, du 26 au 28 juin 2006, durant la Semaine de la Connaissance. Cet atelier, intitulé « Indexation et connaissances en sciences humaines[1] », rassemblera des chercheurs d'une dizaine d'équipe pour approfondir les enjeux liés au développement de ce type de bases de connaissances. Les présidentes en sont Sylvie Calabretto (LISI) du projet Porphyry et Nathalie Aussenac-Gilles (IRIT) pour Arkeotek. Nous en indiquons la grille de travail élaborée pour mener à bien des échanges fructueux. Nous présentons les participants choisis et les retombées scientifiques attendues.

[1] http://www;sdc2006.org/icsh.html

Table des matières

Résumé 3

Table des matières 4

1 Introduction 5

2 Motivations et enjeux du projet Arkeotek 6
2.1 La recherche de connaissances scientifiques au sein d'archives 6
2.2 L'archivage documentaire en sciences humaines 6
2.3 Approches et paradigmes retenus 7
2.4 Archivage et modèles de connaissances 8
2.5 Résultats établis par le projet initial 8

3 Vers un atelier international sur l'indexation de ressources en sciences humaines 9
3.1 Objectifs scientifiques 9
3.2 Méthodologie 9
3.3 Calendrier 10

4 Projets proches de Arkeotek 12
4.1 L'ontologie CRM et le projet CIDOM 12
4.2 Porphyry 12
4.3 Achemenet 13
4.4 Coliscience 14
4.5 HyperNietzche 14
4.6 Le projet Hypertopic 14

5 L'atelier « Indexation et Connaissances en Sciences Humaines » 15
5.1 Choix de s'intégrer à la Semaine de la Connaissance 15
5.2 Participants invités 15
5.3 Grille de questions 16

6 Bilan et perspectives 18
6.1 Travail effectué et avancées 18
6.2 Perspectives 18

7 Publications produites dans le cadre des projets Arkeotek et valorisation 19

8 Références 20

9 Annexes 21
9.1 Bilan financier du projet financé dans le cadre de l'appel « valorisation » 21
9.2 Appel à participation à l'atelier « Indexation et connaissances pour les Sciences Humaines » 21

1 Introduction

Avec l'irruption de l'édition électronique, l'archivage des données scientifiques peut enfin être pensé dans la continuité du travail d'édition des constructions théoriques. La question de l'indexation des données et les paradigmes de consultation sont dès lors profondément modifiés. Pour tirer parti de cette évolution conceptuelle et pratique, les contributions respectives des techniques d'ingénierie de la connaissance et de recherche d'information au sein de documents structurés sont déterminantes et doivent être coordonnées.

Depuis un premier projet mené dans le cadre de l'appel d'offre « archive » (juillet 2003), nous examinons l'utilisation d'outils d'ingénierie des connaissances en association avec des approches innovantes de recherche d'information, tenant compte de la structure des documents. Nous réalisons un prototype de recherche d'informations scientifiques qui permette de définir des vues dynamiques, synthétiques et hiérarchisées, sur des documents et données scientifiques regroupés au sein d'une base de connaissances. Un examen de l'état de l'art nous a permis de repérer des travaux abordant cette problématique de manière plus ou moins proche, comme les projets Porphyry ou CIDOC-CRM qui traitent eux aussi l'archivage et la recherche de connaissances scientifiques dans le domaine de l'archéologie.

Les documents de la base de connaissances sont préalablement structurés selon une méthode originale inspirée des principes du logicisme, puis édités sous forme électronique à l'aide d'un outil d'édition multimédia ad hoc. Plusieurs thèses et articles scientifiques déjà publiés selon ce format forment un point de départ pour l'expérimentation. Une ontologie des domaines scientifiques concernés par ces documents a été construite. Utilisée pour caractériser la sémantique des paragraphes des documents structurés, elle sert d'appui aux procédures d'interprétation des requêtes des utilisateurs et de production documents y répondant. Compte tenu du format original de représentation des documents, la recherche d'information et l'édition des résultats peut alors s'appuyer à la fois sur le contenu, la sémantique des documents, et sur leur structure, révélatrice de l'argumentation scientifique.

Une suite naturelle de ce travail a été de le confronter précisément aux autres expériences de ce type, non seulement les projets Prophyry et CIDOM-CRM, mais aussi les projets consacrés à la recherche d'informations scientifiques en général ou à l'exploration de documents structurés. Le projet dont nous faisons ici le bilan consiste donc à organiser un atelier de travail européen sur plusieurs jours, lieu d'échange approfondi de nos approches et de véritables expérimentations à partir des logiciels disponibles, des corpus de chacun et du format SCD. Cet atelier est envisagé non seulement comme une valorisation scientifique de qualité du travail réalisé, mais aussi comme un moyen de repérer de possibles collaborations en vue d'adapter nos propositions à d'autres domaines ou divers formats de structuration de documents.

Ce rapport présente les éléments préparatoires à l'atelier international que nous organisons en collaboration avec le projet Porphyry. Il se tiendra du 26 au 28 juin 2006, durant la Semaine de la Connaissance. Cet atelier, intitulé « Indexation et connaissances en sciences humaines », rassemblera des chercheurs d'une dizaine d'équipes pour approfondir les enjeux liés au développement de ce type de bases de connaissances. Après avoir rappelé les problématiques centrales d'Arkeotek, qui ont orienté la thématique de l'atelier, nous présentons la démarche retenue pour mettre en place cet atelier. Nous en indiquons la grille de travail élaborée pour mener à bien des échanges fructueux durant l'atelier. Nous présentons les participants choisis et les retombées scientifiques attendues.

2 Motivations et enjeux du projet Arkeotek

Nous récapitulons ici les fondements et problématiques du projet Arkeotek dans le cadre de l'appel d'offre « archives » de 2003, car elles ont orienté la thématique de l'atelier organisé autant que le choix des participants.

2.1 La recherche de connaissances scientifiques au sein d'archives

La problématique qui intéresse le projet Arkeotek dans le cadre de l'appel Archive 2003 est celle de la recherche d'information, plus particulièrement de connaissances scientifiques (en lien avec les raisonnements scientifiques), au sein de collections documentaires structurées et annotées par des connaissances. Cette problématique rejoint plusieurs communautés scientifiques :

- l'ingénierie documentaire et particulièrement la structuration documentaire selon des langages comme XML ;
- l'ingénierie des connaissances et précisément les travaux sur les ontologies, utilisées ici comme des index sémantiques permettant d'assouplir l'association entre requêtes des utilisateurs et parties de documents y répondant ;
- la recherche d'information et surtout les recherches récentes sur les recherches au sein de documents structurés ;
- enfin, et ce n'est pas négligeable, la ou les communautés scientifiques des domaines d'où sont tirés les documents ; en effet, mettre au point ce type de système ne peut se faire sans la participation active des spécialistes du domaine, de leur maîtrise des type de raisonnements mis en œuvre, des concepts du domaine ou encore des types de recherches effectuées.

L'enjeu du projet touche donc l'archivage et l'accès aux savoirs scientifiques sous forme de documents, textes ou images, numérisés, structurés et dont le contenu puisse être caractérisé. Ce travail est mené à bien depuis 2 ans au sein du projet Arkeotek, qui aborde ainsi trois facettes des aspects sémantiques de l'archivage scientifique. Il vise en effet la définition de solutions informatiques favorisant l'archivage de données pertinentes, leur recherche en fonction de points de vue d'argumentation et enfin, leur restitution ou publication auprès de la communauté des chercheurs. Pour cela, les objectifs du projet sont à la fois de favoriser un archivage scientifique, qui s'appuie sur les connaissances relatives aux contenus archivés, et de simplifier l'accès à ces contenus en favorisant l'expression de requêtes liées aux raisonnements et analyses sur ces contenus. Dans le cadre du projet, les connaissances scientifiques concernées touchent au domaine de l'archéologie des techniques. Il s'agit de documents rédigés, de textes, mais aussi de photos, de données chiffrées ou autres références aux objets archéologiques eux-mêmes.

2.2 L'archivage documentaire en sciences humaines

A l'heure actuelle, en Sciences Humaines, les projets d'archivage ont été essentiellement documentaires. Ainsi, des campagnes de numérisation sont régulièrement mises en œuvre pour archiver les données scientifiques. Toutefois, le plus souvent, les chercheurs ne se prêtent pas à ses campagnes, se trouvant devant une difficulté réelle : celle de sélectionner et d'indexer les données à archiver (ex. : immense interrogation devant les mètres linéaires de classeurs de diapositives prises au cours d'années de carrière), sans parler du temps que

l'archivage impose et que les chercheurs ne sont pas prêts à donner dès lors que cet archivage s'inscrit dans une perspective documentaire et non scientifique (publication de leurs travaux).

Ces difficultés de publications des données de recherche et d'archivage posent deux problèmes majeurs : la difficulté d'accès aux données empiriques d'une part, les difficultés de transmission et de capitalisation des connaissances, et donc les progrès des différentes disciplines d'autre part.

Dans le cadre de ce projet, l'archivage scientifique considéré est celui d'écrits (publications scientifiques à caractère monographique ou articles scientifiques) et de données (dessins et relevés, photographies, films, gros volumes de données alpha-numériques) dans le domaine de l'archéologie des techniques. Afin de constituer de véritables bases de connaissances relevant les enjeux épistémologiques et pratiques cités ci-dessus, ces documents scientifiques sont structurés selon les principes de l'analyse logiciste de l'argumentation scientifique en archéologie, puis transformé en document électronique structuré selon un format de représentation, le format SCD. Cette structuration des documents répond en partie aux exigences d'un archivage scientifique permettant la capitalisation des connaissances au sens où nous venons de le définir.

2.3 Approches et paradigmes retenus

Le projet ici présenté a pour ambition de répondre à deux besoins clairement exprimés par la communauté scientifique : 1) l'archivage des données scientifiques ; ce besoin, bien qu'identifié, conduit au mieux à la constitution de bases de données documentaires dont nous analyserons plus loin les limites ; 2) des outils de requête pertinents pour la recherche de résultats scientifiques parmi les nombreuses publications disponibles ; les difficultés actuelles qu'elle pose ont conduit à développer de puissants outils de recherche. Mais ces derniers se trouvent limités dans leurs performances par la forme même des documents scientifiques qui présentent des défauts de structuration face aux impératifs des traitements automatiques.

Face à ce double besoin et les limites des solutions actuellement proposées, notre projet propose qu'en amont les documents soient structurés de telle manière d'une part à intégrer l'ensemble des données mobilisées par le chercheur et ainsi à les archiver selon des considérations scientifiques, d'autre part à développer des outils d'ingénierie linguistique adéquats qui permettent d'offrir des vues dynamiques, synthétiques et hiérarchisées de l'ensemble des documents pertinents pour une requête. La structuration des documents respecte le format d'édition scientifique SCD, format multimédia opérationnel qui s'inspire du logicisme et qui met en scène des documents auparavant ré-écrits en fonction des opérations logico-discursives menées par l'auteur. Les outils d'ingénierie linguistiques sont ceux de la méthode TERMINAE de construction d'ontologie à partir de textes.

Les principes retenus en matière de consultation, en particulier la volonté de favoriser la navigation au sein de la structure des documents et la mise en évidence de « vues » sur ces documents, définies par l'utilisateur en fonction des connaissances recherchées, suppose d'exploiter la structuration des documents au format SCD autant que les index dans la recherche d'information. Cette démarche originale commence à se développer avec la généralisation de l'utilisation du format XML pour la structuration des documents. Du point de vue des systèmes d'information et de connaissances, la problématique concerne donc l'utilisation des ontologies et de la structuration XML pour une recherche d'information « sémantique ». Du point de vue de la consultation et de l'interface, la problématique est celle d'une gestion dynamique des résultats restitués à l'issue d'une recherche.

2.4 Archivage et modèles de connaissances

Ce projet touche des enjeux de société qui vont bien au delà du seul domaine d'application retenu. En initiant ce projet, nous avions identifié les enjeux suivants :

- Montrer les atouts d'une véritable « archivage scientifique » par rapport à un archivage documentaire qui se limiterait aux documents seuls. Par archivage scientifique, nous entendons le fait d'associer des éléments de connaissances permettant de mieux retrouver et exploiter les documents au sein des collections.
- L'utilisation des ontologies en recherche d'information ouvre des perspectives prometteuses pour reformuler des requêtes utilisateur, pour accéder à des données hétérogènes ou encore pour l'indexation sémantique de documents. L'enjeu de société dans le cas de projets largement ouverts et de mise à disposition de connaissances est la légitimité de ces ontologies par rapport au domaine concerné, ainsi que leur adéquation aux collections à consulter.
- Définir de nouvelles pistes pour des consultations moins rigides de ces documents, autorisant de nouveaux points de vue mais aussi la prise en compte d'une logique de lecteur cherchant à recouper des informations d'une manière dissociée des plans de rédaction des auteurs. Cette reconstruction dynamique de nouveaux documents à partir de fragments d'origine rejoint la problématique générale de la consultation de documents numériques structurés.
- Enfin, associer des connaissances de types ontologique ou terminologique à une base documentaire scientifique soulève des questions relatives au point de vue sur le domaine véhiculé par cette ontologie. L'enjeux est ici de trouver un compromis entre les divergences d'opinion ou d'analyse d'un domaine d'une part, et les nécessités de description homogène, voire standardisée, de collections.

Pour un développement complet de ces enjeux, nous renvoyons au rapport final du projet commencé en juillet 2003, remis au programme société de l'information en janvier 2006.

2.5 Résultats établis par le projet initial

Le programme Arkeotek a été envisagé sur 2 ans de septembre 2003 à septembre 2005. Les étapes initialement prévues pour mener à bien ces objectifs correspondent à (1) la définition et l'analyse d'un corpus de référence, (2) au développement d'une ontologie avec une composante terminologique à partir de la terminologie du corpus et de connaissances sur le domaine mais aussi sur le format SCD, (3) à la mise au point et l'évaluation d'un système de recherche d'information dynamique basé sur l'ontologie et la structuration SCD, (4) la définition d'un démonstrateur visualisant les résultats d'une recherche.

Enfin, les travaux en cours contribuent tous à la valorisation du site Arkeotek et de l'ensemble base de connaissances et système de consultation. Les travaux en cours portent sur l'expérimentation du système de consultation, l'enrichissement de l'ontologie et l'ajout de nouveaux documents à la base de connaissances accessible sur le site Arkeotek.org. Cette valorisation prend aussi la forme d'un atelier international co-organisé avec le projet Porphyry, dont les objectifs sont relativement proches, puisqu'il s'agit de permettre l'annotation de corpus de documents de travail des archéologues pour en faciliter l'exploitation

3 Vers un atelier international sur l'indexation de ressources en sciences humaines

3.1 Objectifs scientifiques

Le projet Arkeotek s'inscrit dans une logique de corpus internationaux. Il inclut actuellement plusieurs pays européens qui estiment nécessaire la normalisation des textes scientifiques en SHS, ceci pour un meilleur partage des connaissances et une mutualisation d'instruments de capitalisation et d'interrogation de ces connaissances. La normalisation des textes scientifiques en SHS représente un véritable changement dans nos modes de pensée et nos habitudes d'écriture. Ce changement peut être comparé à celui que connurent les sciences naturelles au XVIIème siècle.

Une poursuite naturelle de ce travail serait de le confronter précisément aux autres expériences de ce type, non seulement les projets Prophyry et CIDOM-CRM, mais aussi les projets consacrés à la recherche d'informations scientifiques en général ou à l'exploration de documents structurés. Nous projet vise l'organisation d'un atelier de travail européen sur plusieurs jours, lieu d'échange approfondi de nos approches et de véritables expérimentations à partir des logiciels disponibles, des corpus de chacun et du format SCD. Il constituera non seulement une valorisation scientifique de qualité du travail réalisé, mais aussi le moyen de déboucher sur de possibles collaborations en vue d'adapter nos propositions à d'autres domaines ou formats de structuration de documents.

A la demande du programme Société de l'Information, nous organisons cet atelier international conjointement avec les membres du projet Porphyry, dirigé par A. Iacovella (ENS, Lyon). Cette collaboration est tout à fait naturelle vue l'apparente proximité scientifique des deux projets. Sylvie Calabretto et Nathalie Aussenac-Gilles sont responsables de l'atelier et sont membre à ce titre du comité de pilotage de la semaine.

3.2 Méthodologie

La méthode retenue pour organiser cet atelier a été de privilégier des contacts scientifiques de qualité et approfondis, permettant d'aller jusqu'à une bonne compréhension réciproque des approches, choix scientifiques et systèmes développés par chacun. Parvenir à un réel échange d'expérience suppose plusieurs pré-requis :

- envisager toutes les facettes des problèmes qui nous intéressent
- bien cibler la communauté de chercheurs concernés
- prendre le temps d'échanger suffisamment d'informations avant de se rencontrer
- établir un partenariat avec 2 ou 3 équipes ciblées pour leur proximité (de domaine d'application par exemple) de manière à diversifier les points de vue y compris dans l'organisation.

Le type de rencontre scientifique envisagé est un atelier de travail au niveau européen. La dimension internationale s'impose pour accéder aux travaux et résultats des projets européens existants. Elle permet également d'aborder des problèmes comme la gestion du multilinguisme dans les ontologies et les ressources documentaires. Le format de l'atelier sera précisé et cela fera partie des premières tâches du projet. Cependant, nous retenons les points forts suivants :

- restreindre le nombre de participants à une trentaine ou quarantaine de personnes ;

- proposer une durée suffisante pour que chacun ait le temps de présenter son travail ;
- proposer une grille de travail commune et précise qui pousse chaque équipe dans une réflexion approfondie sur sa problématique, ses choix fondamentaux, son approche ;
- prévoir des temps d'exposé et des temps de travail collectif autour de questions scientifiques communes ou d'expérimentation à l'aide des systèmes présentés.

Enfin, une publication des conclusions et expérimentations est prévue dans une premier temps par une mise à dispositions des différentes contributions sur le site de l'atelier et sur le site d'Arkeotek. Dans un deuxième temps, nous visons la rédaction, à partir de ces contributions, d'un document collectif sous forme d'article de revue scientifique.

3.3 Calendrier

Ce nouveau projet a commencé en février 2005 et se terminera en juin 2006 avec la fin de l'atelier puis la rédaction d'un article de synthèse. Le programme initial prévoyait la tenue de l'atelier international en janvier 2006. Cet atelier se tient relativement tard par rapport au calendrier prévu d'une part à cause de la nécessaire coordination avec le projet Porphyry, d'autre part afin de progresser dans l'avancement du prototype du système de consultation d'Arkeotek.

En effet, le prototype du système de consultation selon des vues dynamiques devait être opérationnel en août 2005. Or il nous semblait indispensable de pouvoir l'utiliser pour évaluer l'ontologie et mener des expérimentations sur de nouveaux corpus avant de lancer cet atelier de travail. Ce travail est en cours depuis début 2006 pour disposer d'une version opérationnelle et pertinente du logiciel au moment où se tiendra l'atelier de valorisation.

Les tâches réalisées en vue de monter cet atelier sont les suivantes :

Avril 2005 :

- identification d'équipes travaillant sur des projets soulevant la problématique de l'archivage, indexation, recherche d'information pour des bases de connaissances scientifiques.
- Prise de contact en priorité avec les équipes du projet Porphyry, partenaire dans l'organisation de l'atelier à la demande du programme Société de l'information. Rencontre entre A. Besnel, I. Iacovella, V. Roux, S. Calabretto, P. Blasco et N. Aussenac-Gilles.

Juin 2005

- Définition plus précise du contenu, des modalités de participation, du déroulement envisagé, des traces laissées (actes CD et web, puis publication collective après l'atelier).
- Choix d'intégrer l'atelier au sein de la Semaine de la Connaissance 2006 à Nantes

Oct. 2005

- Production d'une invitation à participation.
- Mise en place du site web

Déc. 2005

- Sélection des participants et des projets ou logiciels qui seront exposés et expérimentés ; définition d'un protocole d'expérimentation ou évaluation des différents travaux.

Avril 2006

- Envoie de leur contributions par les participants

Mai 2006

- mise en forme des contributions et premiers échanges

Juin 2006

- Animation de l'atelier et définition de perspectives expérimentales et scientifiques avec les participants.
- Expérimentations sur des outils ou documents ou ressources échangées après l'atelier
- Publication des conclusions et démonstrations éventuelles d'outils sur le site Arkeotek

4 Projets proches de Arkeotek

Un approfondissement bibliographique a permis d'identifier plusieurs projets faisant appel aux ontologies dans le domaine de l'archéologie d'une part, de faire le point sur les approches actuelles en matière de recherche d'information dans des documents structurés, principalement à l'aide du langage XML et enfin de mesurer l'état de l'art en matière de restitution des résultats des requêtes et de vues dynamiques.

Dans les premiers projets utilisant des ontologies en archéologie, nous avons cherché d'éventuelles ontologies en vue de les réutiliser ainsi que des résultats sur le type d'ontologie pertinent pour faciliter la recherche de données ou textes archéologiques au sein de collections. Les projets Porphyry et CIDOC / CRM sont de bonnes illustrations de ces recherches.

4.1 L'ontologie CRM et le projet CIDOM

Le projet européen CIDOC soutenu par l'ICOM (International Council of Museums) vise une meilleure gestion des collections, archives et produits scientifiques ou administratifs relatifs au patrimoine artistique et culturel (CIDOC/CRM, 2002). Il s'agit de rendre plus facilement accessible l'information et les données culturelles, de mieux les archiver, les préserver, les retrouver, les confronter, etc. que ce soit pour des analyses scientifiques ou pour les présenter au public. L'idée initiale est de mettre en place un modèle de données standard qui facilite la description des objets et informations culturels, qui soit exploitable par des systèmes informatiques et soit suffisamment riche pour rendre compte des diversités d'analyse et d'interprétation.

Un modèle a donc été mis au point, le CRM, pour Conceptual Reference Model, qui spécifie sous forme d'ontologie formelle un certain nombre de concepts et de relations entre eux jugés utiles pour la description d'entités culturelles et des relations qu'elles entretiennent entre elles (images, textes, lieux historiques, objets archéologiques mais aussi textes à propos d'images, d'événements, etc.). Cette ontologie formelle, écrite en TELOS mais disponible aussi en RDFs (format compatible avec XML), permet à la fois d'associer des méta-données aux objets ou aux documents (donc de les indexer), mais aussi d'effectuer des recherches dans une base d'objets ainsi indexée.

Cette ontologie CRM est supposée couvrir les œuvres d'arts autant que les objets archéologiques, leur situation dans le temps et l'espace ou encore les interprétation qui en sont faites, la gestion de leur propriété, etc. La partie relative à la description des périodes temporelles, des zones géographiques et des objets artistiques ou socioculturels nous a semblé pouvoir servir de référence pour établir l'ontologie Arkeotek. Le fait de considérer l'ontologie d'Arkeotek comme une spécialisation de l'ontologie CRM permettrait de se situer par rapport à un standard et de s'y conformer en partie, de bénéficier des travaux déjà effectués pour parvenir à CRM en matière d'analyse du domaine, et de s'appuyer sur une représentation formelle éprouvée. Ces perspective a donc été envisagée.

4.2 Porphyry

Le projet Porphyry (http://www.prophyry.org) comprend plusieurs aspects. Retenons-en la première plate-forme logicielle présentée. L'approche part du principe que « dans les domaines d'expertise, chacun apporte son modèle personnel vis-à-vis d'un corpus documentaire de référence » et que « la confrontation des points de vue stimule une évolution

constante des modèles. Né d'une dynamique interdisciplinaire (archéologie, informatique...), le projet Porphyry se propose donc d'instrumenter ce travail d'expertise. Il est fondé sur l'idée qu'une telle pratique "laisse des traces" : sélections de fragments, documents ajoutés au corpus, organisation de la bibliographie, notes de lecture, etc. A condition que ces traces soient "captées", leur auteur peut désormais les utiliser pour accéder au corpus, ainsi que les partager avec d'autres experts. En somme, puisque le corpus se structure de pair avec le savoir de l'expert au fil de ses changements de problématique et de point de vue, nous proposons un atelier de manipulation de corpus documentaires afin d'assister un travail humain de construction de sens ».

Le projet présente donc quelques particularités fortes.

Il fait l'hypothèse de l'existence d'un corpus documentaire partagé, de référence. Le projet regarde dès lors le travail de consultation et d'expertise, comme un exercice de sélection, d'annotation et in fine d'extension de ce corpus documentaire. Chaque « point de vue » est un cheminement possible dans l'espace des documents (données, annotations et constructions théoriques ne sont pas distinguées mais au contraire regardées comme des documents traces).

Parce que chaque objet documentaire peut se retrouver à « l'embranchement de parcours appartenant à différents points de vue », cette approche se démarque vigoureusement des approches de types thésaurus dans l'indexation des documents. En effet, plutôt que de contraindre les experts du domaine à se plier à une représentation unique intégrant tous les points de vue existants (et forcément divergents) sur le domaine, le système préfère en favoriser la juxtaposition. Cette approche est orthogonale à une approche ontologique, qui tendrait à proposer un point de vue unique sur le domaine. Ce choix se justifie aussi par la couverture assez large des thèmes abordés dans les corpus.

Une des limites de ce choix est bien sûr qu'un point de vue est difficilement accessible en dehors de son auteur. En effet, l'interface de consultation ne permettant d'avoir une vue globale sur l'ensemble des points de vue, l'utilisateur ne sait pas quels sont les concepts selon lesquels il peut interroger la base. Un autre inconvénient est que les différents points de vue sont parfois contradictoires. Pour gérer ce problème, des fonctions de comparaison sont actuellement développées en vue de détecter ces contradictions et de les signaler.

Ces choix reviennent à privilégier une consultation du fond par les auteurs et ne convient donc pas pour notre projet, qui cherche à privilégier l'accès au contenu des documents par l'ensemble de la communauté scientifique concernée. Cependant, la représentation des points de vue est un bon indicateur du type de connaissances que peut contenir l'ontologie qui nous intéresse.

4.3 Achemenet

http://www.achemenet.com/

Ce site facilite la recherche de données dans le domaine de l'assyriologie. Ces données sont des documents textuels ou des photographies d'objets relatifs à la culture des achéménides. Le site et l'application associés cherchent à favoriser le partage de leurs sources et de leurs documents scientifiques par les chercheurs, de manière à favoriser une dynamique de la recherche dans le domaine. Plusieurs arborescences de mots-clés servent à structurer la mise en ligne de ces textes en archéologie. Chaque arborescence correspond à un point de vue sur le domaine, et reflète une sorte de modèle de connaissances qui donne des points d'entrée dans la base documentaire.

4.4 Coliscience

http://www.colisciences.net

Le projet Coliscience représente une expérience de mise à disposition de documents en vue de leur exploitation scientifique par toute communauté de chercheurs, et ce pour des objectifs non prédéfinis. L'effort porte à la fois sur la numérisation de documents d'archive, et la facilité de recherche d'informations précises au sein de ces documents. Les textes appartiennent à la littérature scientifique du 19e S., en l'occurrence l'œuvre de Claude Bernard.

4.5 HyperNietzche

http://www.hypernietzsche.org/

Dans ce projet, les documents mis en ligne concernent non seulement les écrits de Nietzche mais aussi des écrits sur l'œuvre de Nietzche, des interprétations possibles de son œuvre par d'autres philosophes ou des chercheurs. Chaque navigation au sein du corpus produit une interprétation particulière des textes de Nietzsche. L'ensemble de ces interprétations forme effectivement un "parcours de recherche".

Comme le projet Porphyry, HyperNietzche s'intéresse à la manière dont le scientifique se sert de ses sources pour produire un résultat. Dans ce projet, le parcours de recherche produit par un utilisateur en fouillant la base de données devient à son tour une donnée épistémologique intéressant la validité du résultat.

4.6 Le projet Hypertopic

http://tech-web-n2.utt.fr/

Le modèle HyperTopic est adapté notamment à toutes les utilisations à base de traces, au sens large du terme, et permet la construction collective et permanente de Cartes de Thèmes multi-points de vue, support de ces traces. En se basant sur ce modèle, les applications peuvent utiliser, réutiliser ou échanger plus facilement des Cartes de Thèmes multi-points de vue, dans l'esprit du web socio sémantique qui ne vise pas des « ontologies de domaine », formellement univoques et imposées une fois pour toutes aux acteurs."

5 L'atelier « Indexation et Connaissances en Sciences Humaines »

5.1 Choix de s'intégrer à la Semaine de la Connaissance

Nous avons choisi d'organiser cet atelier au sein de la Semaine de la Connaissance, regroupement de manifestations scientifiques portant sur l'étude des connaissances dans le cadre des nouvelles technologies de l'information et de la communication. Ainsi, nous pouvons bénéficier de l'organisation matérielle de cet événement, et surtout organiser l'atelier en deux temps :

- deux journée réservées à un travail approfondi d'échange entre les équipes participantes ; ces journées ne sont pas fermées mais le format retenu privilégie un travail scientifique intéressant essentiellement les équipes invitées ; d'autres chercheurs peuvent toutefois participer comme auditeurs ;
- une demi-journée « ouverte » donne l'opportunité de restituer une synthèse « à chaud » et de la présenter à des discutants choisis pour leur esprit critique sur le sujet ; nous espérons ainsi donner une nouvelle dimension à la confrontation des différents travaux, et poursuivre le débat pluridisciplinaire qui est au cœur de la production de sites et ressources numériques au service des sciences humaines.

L'atelier aura donc une dimension internationale, sera de langue anglaise, et constituera un des événements de la Semaine de Connaissance à Nantes. Il aura lieu du 26 au 28 juin 2006.

Le site web de l'atelier se trouve à l'url suivant :

http://www.sdc2006.org/ICSH.html pour la version française et

http://www.sdc2006.org/IKHS.html pour la version anglaise.

Les deux co-responsables de l'atelier sont Sylvie Calabretto (LISI) pour le projet Porphyry et Nathalie Aussenac-Gilles (IRIT) pour le projet Arkeotek.

5.2 Participants invités

Les projets suivants ont été invités :

CIDOC - CRM (http://cidoc.ics.forth.gr/) : Martin Doerr (Amsterdam)

MICADO (http://web.upmf-grenoble.fr/adest/seminaires/sem5mai/gallezot/Gallezot2.pdf) : Philippe Bessières (INRA, Jouy en Josas) Normes et standards pour la procédure de traitement du document numérique en biologie moléculaire.

ACHEMENET (http://www.achemenet.com/) : Pierre Briant (Collège de France), José Paumard (Paris-XIII)

HYPERNIETZCHE : http://www.hypernietzsche.org/

ARMADILLO (AHR : Armadillo for History Repositories, utilisation de Armadillo pour indexer des documents en sciences humaines) (http://nlp.shef.ac.uk/wig/) : Mark Greengrass (Sheffield, UK)

COLISCIENCE : http://www.colisciences.net Jean Vignaud (CNRS, Paris)

HYPERTOPIC et AGORAE (http://tech-web-n2.utt.fr/) : Aurélien Besnel et Manuel Zacklad (TechCico, UTT, Troyes)

Les discutants invités sont les suivants :

D. BOURCIER, spécialiste en droit et sciences cognitives (CNRS, Paris), B. BACHIMONT, spécialiste en philosophie, ingénierie documentaire et document multi-média (INA, UTC), W. STOCKOWSKI, anthropologue (EHSS, Paris).

5.3 Grille de travail

Cette grille de questions est le support des réflexions préalables demandées aux équipes participantes. Elle est la trame du document préparatoire de 4 à 5 pages demandé à chaque participant. Elle a pour objectif également de permette à chacun de connaître précisément les différents projets participant à l'atelier. Enfin, les journées s'organiseront selon cette grille : les différentes sessions correspondent à chacune des parties de cette grille. Au cours des sessions, chacun des participants présentera les réponses et points de vue de son projet.

1) Présentation rapide des approches :

- logiciel proposé : fonctionnalités, utilisateurs ciblés

- objectifs visés par le projet pour le chercheur en SHS (se situer par rapport à 1 des points suivants) :

1. retrouver le "savoir" que mobilise un scientifique pour énoncer un résultat ;

2. étudier – éditer la manière dont le scientifique se sert de ses sources pour produire un résultat (le parcours de recherche=une donnée épistémologique intéressant la validité du résultat) ;

3. partager des sources et des documents scientifiques de manière à favoriser une dynamique de la recherche ;

4. aider à la recherche sur les données primaires.

- données manipulées : sources brutes (sources primaires) versus sources retravaillées pour être indexées (sources secondaires)

2) Présentation détaillée des approches :

- Nature des modèles :

- type de contenu,
- modèle de données,
- degré de formalisation,
- ambition : création de connaissances, élaboration de consensus, support au dialogue entre courants scientifiques, etc.

- Nature des documents :

- descriptions factuelles versus raisonnements, textes argumentatifs
- degré de structuration des textes : linéaires vs structurés

- Mode de construction des modèles

- quelles sources de connaissances sont-elles exploitées ?
- qui maîtrise le contenu des modèles ?
- cycle de vie, processus de maintenance des modèles

- Liens entre modèles et documents (indexation ? annotation ?) ; qui les fixe, dans quel objectif ?

3) Aspects méthodologiques, place de l'interdisciplinarité

- Comment les approches précédentes ont-elles été définies ?

- Quelle est la place de la réflexion des sciences humaines sur leur patrimoine de données et leur accès aux données ?

- Quelle est la part de l'influence technologique ?

- En quoi l'interdisciplinarité est-elle un enjeux fondamental pour mieux étudier la question de l'accès au contenu ? Comment la mettre en place ?

4) Démonstrations, atelier d'utilisation d'outils

- protocole d'utilisation des outils ; des résultats

- En quoi les outils ou environnements proposés facilitent-ils un meilleur accès à des ressources ou des collections utiles à l'activité scientifique en sciences humaines : données, connaissances et résultats sur ces données ?

- En quoi instrumentalisent-ils le travail du chercheur ? comment ?

5) Comparaison des différents travaux (modèles et logiciels) sur une question particulière :

- devant une question de représentation de connaissances en matière de périodisation, que propose chaque approche ?

- devant une question de divergences d'interprétations, que propose chaque approche ?

6 Bilan et perspectives

6.1 Travail effectué et avancées

Au moment de rédiger ce rapport, l'atelier est mis en place. Plus de la moitié des projets contactés ont confirmé leur participation. La thématique retenue s'avère tout à fait d'actualité. En effet, l'indexation est une question largement abordée du point de vue technique en recherche d'information. L'ingénierie des connaissances la traite sous un angle un peu différent et s'y intéresse de plus en plus dans le contexte du Web Sémantique. Or la question des contenus, de l'adéquation avec les pratiques des utilisateurs d'une part, des communautés produisant les fonds documentaires d'autre part, reste peu abordée. Il est donc tout à fait opportun de la mettre en avant car elle conditionne (tout autant que la qualité technologique) l'utilisant réelle de ces systèmes de recherche d'information et des serveurs de ressources développés. Nous avons donc identifié une thématique originale qui soulève des questions fondamentales en lien avec la gestion des connaissances et la numérisation des documents.

Ce projet d'atelier international a permis d'autres avancées à nos équipes. En particulier, en prenant contact avec des participants, nous avons établis des relations avec la plupart des équipes européennes intéressées par des approches pluridisciplinaires pour la gestion des connaissances en sciences humaines. Il s'agit là d'un moyen très pertinent pour mieux maîtriser notre position par rapport à l'état de l'art.

Parmi les quelques difficultés liées à la mise en place de cet atelier, mentionnons l'ajout de temps que requiert la coordination avec les membres du projet Porphyry et le temps plus long que prévu requis par les prises de contact avec d'autres projets. En situant l'atelier pendant la semaine de la connaissance, nous avons retardé volontairement notre calendrier, ce qui nous permet de développer et évaluer un prototype plus performant.

6.2 Perspectives

Plusieurs suites à donner à cet atelier sont envisagées si jamais l'atelier est un succès :

- Une première piste est de le renouveler si une communauté scientifique se dégage et si des questions méritent d'être approfondies.
- A minima, nous souhaitons entretenir les contacts avec les équipes participantes, favoriser les échanges, de manière à diversifier nos références et les tenir à jour.
- Enfin, ces contacts pourraient déboucher soir sur collaborations bi-partites, soit sur la mise au point d'un réseau européen ou de projets internationaux.

7 Publications produites dans le cadre des projets Arkeotek et valorisation

AUSSENAC-GILLES N., ROUX V, de SAIZIEU B., BLASCO P., Ontologies dédiées à la consultation de documents structures selon un modèle logico-sémantique. In *Actes du colloque de clôture du programme Société de l'Information*. Lyon (F), 19-21 mai 2005. 13-16. 2005.

GARDIN, J-C. et ROUX, V. 2004. The Arkeotek project : a european network of knowledge bases in the archaeology of techniques. *Archeologia e Calcolatori*, 15, 25-40

ROUX, V. avec la participation de Ph. BLASCO 2004. Faciliter la consultation de textes scientifiques. Nouvelles pratiques éditoriales. *Hermès, Critique de la raison numérique*, CNRS éditions, 39, 151-159.

Nouveau volume de la collection Référentiels :

GELBERT, A. *Traditions céramiques et emprunts techniques dans la vallée du Fleuve Sénégal*, Paris, Editions de la MSH / Editions Epistèmes, 101 p. Collection référentiels n°1. Cédérom bilingue inclus.

BOILEAU M.-C., *Production et distribution des céramiques au IIIe millénaire en Syrie du Nord-Est*. Éditions de la MSH / Éditions Épistèmes, Paris. Collection référentiels n°2. 2005. Cédérom bilingue inclus.

Rapports sur le projet Arkeotek :

N. AUSSENAC-GILLES, V. ROUX, P. BLASCO, *Rapport intermédiaire du projet Arkeotek : Utilisation d'ontologies pour la définition de vues dynamiques permettant la lecture par niveaux d'une base de connaissances de textes et de données scientifiques structurés selon le format SC*. Rapport interne IRIT/2004-20-R. Nov. 2004. 27 p.

PUJOS N., *Spécification d'un outil de recherché dans des documents structurés*. Rapport de stage IUP ISI, 3eA. Université Paul Sabatier, Toulouse. Sept. 2004.

BERNADOU P., *Système de recherche sémantique d'information au sein de documents structurés*. Rapport de stage IUP ISI M1. Université Paul Sabatier, Toulouse. Sept. 2005.

CZERNY J., *Système de recherche sémantique d'information au sein de documents structurés*. Rapport de stage IUP ISI M1. Université Paul Sabatier, Toulouse. Sept. 2005.

8 Références

AUSSENAC-GILLES N., BIÉBOW B., SZULMAN N., 2000 Revisiting Ontology Design: a method based on corpus analysis. *Knowledge engineering and knowledge management: methods, models and tools, Proc. of the 12th International Conference on Knowledge Engineering and Knowledge Management*. Juan-Les-Pins (F). Oct 2000. R Dieng and O. Corby (Eds). Lecture Notes in Artificial Intelligence Vol 1937. Berlin: Springer Verlag. 2000. 172-188.

N.AUSSENAC-GILLES, A.CONDAMINES, 2001 : " Entre textes et ontologies formelles : les Bases de Connaissances Terminologiques ". M.Zacklad et M.Grundstein (eds) : Ingénierie et capitalisation des connaissances. Paris : Hermès. 2001, 153-176.

CIDOM/CRM : *Definition of the CIDOC object-oriented Conceptual Reference Model*. Eds N. Crofts, M. Doerr, T. Gill, S. Stead M. Stiff. Version 3.4. Nov. 2002

GARDIN, J.C. 1979. *Une archéologie théorique*. Paris : Hachette.

GARDIN et al. 1987. *Systèmes experts et sciences humaines*, Paris : Editions Eyrolles.

GARDIN, J.C. 1991. *Le calcul et la raison. Essais sur la formalisation du discours savant*.. Paris : Editions de l'EHESS.

GARDIN, J.C. 1998. *Prospections archéologiques en Bactriane (1974-1978), description des sites et notes de synthèse*. Paris : Editions Recherche sur les civilisations.

A. MAEDCHE, S. STAAB, "Semi-automatic Engineering of Ontologies from texts", Conference on Software Engineering and Knowledge Engineering. Chicago (USA), 2000.

ROUX, V. sous presse. Modélisation des constructions scientifiques sur multimédia et transfert des connaissances. In P. GESLIN, D. MAGDA, P. SALEMBIER, C. ALBALADEJO *La transférabilité des connaissances en question(s*). Paris : Editions du CNRS.

ROUX, V. (sous la dir.) 2000. *Cornaline de l'Inde. Des pratiques techniques de Cambay aux techno-systèmes de l'Indus*. Paris : Editions de la MSH, 545 p. Cédérom bilingue inclus.

ROUX, V. ET BLASCO PH.. 2004. Logicisme et format SCD : d'une épistémologie pratique à de nouvelles pratiques éditoriales *Hermès*.

SALAÜN J.M., L'édition entre biens et services. *Interdisciplines : The future of Web publishing*, 2002.

SZULMAN, S., BIÉBOW B., AUSSENAC-GILLES N., 20002. Structuration de Terminologies à l'aide d'outils d'analyse de textes avec TERMINAE. *Traitement Automatique de la Langue (TAL).* Numéro spécial sur le Structuration de Terminologie. Eds A. Nazarenko, T. Hammon. Vol 43, N°1. 103-128. 2002.

WU Z., PALMER M., Verb Semantics and Lexical Selection*, Proceedings of the Annual Meeting of the Association for Computational Linguistics. ACL* 1994

9 Annexes

9.1 Bilan financier du projet financé dans le cadre de l'appel « valorisation »

Crédits Notifiés (HT) 2005	10000
Dépenses	3000
Engagé en 2006	7000
Reste	0

Détails Dépenses 2005-2006		**dépenses**	**engagé**
	dépenses relatives à l'enrichissement du corpus		
10/10/2005	Editions Epistemes	3000	
	Missions relatives à l'organisation de l'atelier		**810**
13/03/2006	N. Aussenac-Gilles		420
06/04/2006	N. Aussenac-Gilles		390
	Inscriptions atelier et séjours		**2590**
26/06/2006	N. Aussenac-Gilles		640
26/06/2006	V. Roux		450
26/06/2006	B. Barthélémy		450
26/06/2006	P. Blasco		450
26/06/2006	A. Reymonet		600
	Missions des conférenciers invités		**1800**
25/06/2006	P. Bessières INRA, Jouy en Josas)		800
25/06/2006	Mark Greengrass (Sheffield)		1000
	Fonctionnement - stages		**1800**
05/05/2006	finalisation de l'atelier Arkeotek		1800
	TOTAL	**3000**	**7000**

9.2 Appel à participation à l'atelier « Indexation et connaissances pour les Sciences Humaines »

9.2.1 Version française

Bonjour,

Dans le cadre de la Semaine de la connaissance, organisée du 26-30 juin 2006 à Nantes (http://sdc2006.org/), les projets Porphyry (www.porphyry.org) et Arkeotek (www.arkeotek.org) organisent un atelier international sur "Indexation et Connaissances en Sciences humaines" du 28 au 30 juin 2006.

Votre projet ... relevant de cette problématique, nous avons le plaisir de vous inviter à y participer, de manière à confronter votre approche à celles des autres participants. Vous trouverez ci-dessous une présentation détaillée des objectifs de cet atelier, ainsi que la grille de questions à laquelle chaque équipe doit répondre au sein d'une communication écrite en anglais à remettre pour le 15 mars 2006.

L'atelier prendra en charge les frais de mission (inscription à la SdC, voyage et séjour) d'un représentant de votre projet. D'autres membres du projet peuvent également participer, mais à leurs frais.

Merci de nous indiquer rapidement si vous acceptez ou non cette invitation.

Nous restons à votre disposition pour toute précision au sujet de cet atelier, ...

Bien sincèrement,

Société de l'Information

...

Objectifs de l'atelier

Depuis la généralisation du support numérique pour conserver et diffuser des données d'étude (images, textes, etc.) mais aussi des analyses scientifiques et des connaissances sur ces données, les chercheurs en sciences humaines et les informaticiens se heurtent à la question de l'accès au contenu. Comment caractériser, représenter des documents (connaissances ?), restituer les spécificités d'un domaine, mais aussi comment donner à différents types d'utilisateurs, les moyens de retrouver des informations précises, données ou raisonnements ? Les techniques utilisées peuvent faire appel à l'indexation, l'annotation, l'archivage et la recherche d'information pour l'aide à la publication, à la formulation et à la diffusion de connaissances scientifiques, de données et de ressources sur ces données. Plusieurs de ces travaux ont débouché sur des logiciels et des plates-formes destinées aux chercheurs en sciences humaines.

L'objectif de cet atelier est de favoriser une meilleure connaissance réciproque des équipes ayant contribué à ces systèmes et de progresser dans les questionnements scientifiques relatifs à ces problématiques. Il s'agit de mesurer la portée de chaque projet, son niveau de développement, de connaître ses applications s'il en existe, d'identifier les modèles et leurs atouts pour caractériser et retrouver le contenu de documents et ressources. A terme, nous souhaitons que ces échanges débouchent sur une publication scientifique dressant un état de l'art, soulignant les enjeux scientifiques pluri-disciplinaires liés à l'accès au contenu et faisant le point des méthodes et des technologies pertinentes pour ce type d'application.

Grille de présentation des travaux des participants

Nous proposons une grille d'analyse des différentes approches qui servira à structurer l'atelier. Les participants devront présenter leurs travaux selon le canevas de cette grille dans un document de participation rédigé en anglais, de 3 à 5 pages, et ce avant le 15 mars 2006. Ce document doit être envoyé à Sylvie Calabretto sylvie.calabretto@insa-lyon.fr, suivant le format suivant : http://www.sdc2006.org/proposition%20article.html

Grille présentée plus haut dans le rapport.

9.2.2 Version anglaise

Goals of the workshop

Since the generalization of the numeric format to store and to exchange study data (images, texts, etc.) as well as scientific analyses and thus knowledge about data, researchers in human sciences and in computer science have started to tackle the question of how to access to the content of scientific documents. For this purpose, how could these documents be represented and characterized? How is domain specific knowledge retrieved? And how can various users be given the means to find out precise information pertaining either to data or reasoning? The techniques used to answer may include indexing, annotation, archive management and information retrieval. They should help to publish, formulate or disseminate scientific knowledge, data and resources about these data. Several software and frameworks developed for researches in human sciences are today available.

This workshop aims at providing an opportunity for the teams that contributed to these systems to meet and know better each other. It will also be a way to progress in the scientific questions related to these topics. One of the motivations is to measure the scope and degree of achievement of each project, to know its applications if any, to identify the models and their strengths that help to characterize and search for the content of documents and resources.

As a final goal, we wish that these discussions will lead to a scientific publication that will draw a state of the art. It will also stress the cross disciplinary scientific challenges related to the access to document content. This paper should finally report the main methods and technologies required for this kind of applications.

Work presentation road-map for the participants

We suggest a road-map so that all the participants present their own approach according to a similar structured way, that will provide guidelines to the workshop. All the participants should present their work according to this roadmap in a paper communication which should be in English, from 3 to 5 pages long. It should be sent before March 15th to Sylvie Calabretto sylvie.calabretto@insa-lyon.fr .

The paper format is available at http://www.sdc2006.org/proposition%20article.html

1) Short presentation of the approach :

- Goal of the project from the point if view of a researcher in Human Sciences. Please try to situate your contribution in one of the following purposes :

  1. To retrieve in documents the "knowledge" that scientists (in human sciences) mobilize in order to establish a result (reasoning underlying scientific results);

  2. To study or edit the way a scientific makes use of sources to produce a result (research steps is considered here as an epistemological data to measure the result validity) ;

  3. To share scientific sources and documents in order to stimulate the dynamic of research ;

  4. To develop search tools to be applied on primary data

- handled data: raw data (primary sources) against sources modified for a better indexing (secondary sources)
- If any software : functionalities, target users ;

2) Detailed presentation of the approach:

- Nature of the models:
- Type of content;
- Data model;
- Degree of formalization;

Ambition: to create some knowledge, to reach a consensus, to support dialogue between scientific streams, …

- Nature of the documents:

Factual descriptions versus reasoning (argumentative texts), degree of structure of these texts : linear vs. structured, …

- Way of designing the models:
  - What are the used knowledge sources used? who is responsible for the model contents?
  - What are the model life cycle? Is there any maintenance process?

- What are the connexions between models and documents (indexing? annotation?) Who is setting these connexions and how? what is the purpose?

3) Methodological aspect, importance of cross-disciplinary research:

- How was the approach defined?
- What is the official position of human scientists concerning their data patrimony and their access to data?
- How much does technology influence this position?
- Is a cross-disciplinary research a major challenge to better study the question of the access to the content of documents? How can it be set?

4) Demonstrations, practical session to test and use the tools:

- protocol to use the tools and its results ;
- Are the tools or frameworks providing an easier access to some resources or collections useful in the researchers' activity in human sciences : data, knowledge and results about these data?
- Are they forming new tools to be used by the researcher? How?

5) Comparison of the different approaches (models and tools) about one particular question:

- what does each approach propose to represent knowledge about dating (period identification)?

- what does each approach propose when dealing with contradictory interpretations ?