



68th IFLA Council and General Conference

August 18-24, 2002

Code Number: 006-122-F
Division Number: IV
Professional Group: Classification et indexation
Joint Meeting with: -
Meeting Number: 122
Simultaneous Interpretation: -

Interopérabilité des accès matière : conclusions du projet HILT (High Level Thesaurus)

Dennis Nicholson

Strathclyde University
Glasgow, Royaume-Uni

Introduction : de HILT Phase I à HILT Phase II

Les questions concernant l'interopérabilité des accès matière exposées dans cet article résultent de deux projets, aujourd'hui appelés HILT phase I et HILT phase II.

HILT phase I (auparavant désigné seulement sous le nom de projet HILT) a fait l'objet d'un rapport en novembre 2001. Il a été mis en place conjointement par le Joint Information Systems Committee (JISC, Comité commun des systèmes d'information) et le Research Support Libraries Programme (RSLP, Programme d'aide à la recherche pour les bibliothèques) et a juste duré une année. On s'attend à ce que le soutien par JISC de HILT phase II soit confirmé en avril 2002.

La deuxième phase durera 12 mois et s'appuiera sur le travail de HILT phase I, ainsi que sur les qualités et l'expérience de l'équipe qui a porté le projet. A partir du consensus réalisé dans la première phase, elle visera à créer un service pilote de mapping terminologique, ou "navigateur terminologique" (*route map*), avec une attention particulière au développement de la Distributed National Electronic Resource (DNER, Ressource électronique nationale répartie) incluant, mais pas nécessairement de façon restrictive, l'enseignement supérieur et la formation continue, en insistant sur les terminologies d'accès matière pour la description au niveau des collections dans l'environnement documentaire (IE) prévu par le JISC.

L'évaluation par l'utilisateur et l'analyse coût-efficacité des divers niveaux de service seront également des caractéristiques du projet.

Le problème

Pour assurer que les utilisateurs de l'IE (l'environnement documentaire prévu) peuvent trouver les ressources appropriées, l'accès par sujets est un des défis principaux que doivent relever le JISC, la DNER, le Resource Discovery Network (RDN, Réseau d'accès aux ressources) et les divers fournisseurs de service à travers les archives, les bibliothèques, les musées, et les services électroniques.

Comme HILT phase I l'a montré, les divers fournisseurs de service emploient une large gamme de langages d'indexation (des systèmes d'accès matière généraux comme les LCSH, le thésaurus de l'Unesco, la DDC (CDD) et l'AAT jusqu'aux spécifiques MeSH) pour répondre à l'exigence de décrire de façon satisfaisante et cohérente leurs ressources pour permettre une récupération précise.

Afin que la recherche par équation et par feuilletage fonctionne de manière cohérente pour les utilisateurs de l'IE, ces systèmes d'accès matière doivent être "mappés" entre eux, peut-être en utilisant une "épine" commune comme la DDC (CDD, Classification décimale Dewey) avec une application internationale et multilingue et la capacité de faciliter le travail réciproque de machine à machine. Et, plus important peut-être encore, les termes présents dans les esprits des différents types d'utilisateurs doivent être "désambiguïsés"¹, puis traduits dans les termes effectivement employés à l'indexation et qu'il faut utiliser pour lancer la recherche dans le groupe des services aptes à répondre à la requête. Le but de HILT phase II est d'établir et d'évaluer un service pilote qui sera l'intermédiaire dans ce processus, comme "service partagé" de la DNER dans l'IE.

HILT Phase I

HILT phase I a démontré que :

- beaucoup de systèmes d'accès matière et de nombreuses pratiques sont en usage dans les services britanniques qui pensent que la recherche par sujet est utile à la fois pour les utilisateurs et les professionnels.
- Il y avait un fort consensus de la part des archives, des services électroniques, du milieu des bibliothèques et des musées en faveur d'un projet pilote plus concret qui, tout en déterminant ses coûts et ses avantages, développerait en réseau un "navigateur terminologique" interactif, réactif à l'utilisateur comme à la machine. Cette "route map" serait la voie d'accès aux terminologies employées par ces institutions et aux rapports d'équivalence entre ces terminologies (projet précédemment désigné dans HILT comme "service pilote de mapping" - voir le diagramme TeRM (*Terminologies Route Map*) en Annexe A pour une description d'ensemble de ce qu'il recouvre et de la façon dont il fonctionnerait).
- Une recherche plus poussée était exigée sur l'efficacité, le niveau et la nature des besoins des utilisateurs, le caractère pratique, les conditions de conception, et les coûts-avantages d'une telle approche avant qu'un engagement à long terme pour un service probablement cher puisse être justifié. Toutes choses qui, comme on l'a déterminé, pourraient mieux être étudiées à travers un projet pilote qui examinerait ces questions.

¹ Désambiguïser = déterminer si l'utilisateur qui saisit "lotus" cherche à obtenir des informations sur la voiture, le logiciel, la fleur, ou un autre des nombreux sens de ce terme

Des informations complémentaires sur HILT phase I peuvent être trouvées sur le site Web de HILT² en général, et plus particulièrement dans le Rapport final HILT³.

HILT Phase II : les objectifs

HILT phase II met en place ce processus dans l'étape du projet pilote, en se concentrant - comme le recommandait l'évaluation de HILT phase I - sur les exigences d'une terminologie et d'un thésaurus au niveau de la collection, mais considérant également la nécessité d'étendre ces exigences en temps opportun aux besoins liés à la récupération au niveau de l'item.

Il utilisera le travail effectué par HILT phase I, et s'appuiera sur les qualités et l'expérience de cette équipe pour mettre en place un service pilote de "navigation terminologique" (service TeRM), semblable à ce qu'avait proposé HILT phase I, et visant à :

- a. Fournir une base expérimentale dans laquelle analyser et établir le service d'accès matière approprié à l'IE du JISC, l'accent étant mis plus particulièrement sur la DNER, le RDN, l'utilisateur, le niveau collection, la compatibilité internationale, ainsi que sur les considérations de l'accès local, régional, national et britannique dans son ensemble.
- b. Faire des recommandations pour un possible service futur, en prenant en compte un ensemble de facteurs, y compris le niveau et la nature des besoins des utilisateurs, le caractère pratique, les conditions de conception, l'efficacité, les fonctionnalités disponibles dans les progiciels commerciaux existants en comparaison avec un produit "maison", et (surtout) les coûts-avantages pour les utilisateurs de l'enseignement supérieur et de la formation continue d'un outil de recherche par sujets reposant principalement sur les besoins au niveau de la collection.

HILT Phase II : les participants

HILT phase II durera 12 mois, et impliquera en gros le même ensemble de participants que HILT phase I, mais avec une participation plus directe de représentants de la Distributed National Electronic Resource (DNER), du Resource Discovery Network (RDN) et de la formation continue.

Plus précisément :

- CDLR, Centre for Digital Library Research (CDLR) de la Strathclyde University, leader
- Représentant de la DNER
- mda (ex- Museums Documentation Association)
- NCA (National Council on Archives)
- NGfL (National Grid for Learning), Écosse
- OCLC (Online Computer Library Center)
- Représentant du RDN
- Représentant de la formation continue
- SLIC (Scottish Library and Information Council)
- SufI (Scottish University for Industry)
- UKOLN (UK Office for Library and Information Networking)
- Des experts en terminologie, Alan Gilchrist et Leonard Will (intervenant extérieur).

² <http://hilt.cdlr.strath.ac.uk/>

³ <http://hilt.cdlr.strath.ac.uk/Reports/FinalReport.html>

Par sa participation au groupe de projet CAIRNS⁴ (qui a utilisé des collections très importantes pour aménager des mini-ensembles), aux projets SCONE et SEED⁵ qui ont contribué à élaborer une base de données des collections multi-sectorielles⁶, ainsi que par sa participation à HILT⁷ phase I, le site leader - CDLR⁸, Centre de recherche pour la bibliothèque numérique de la Strathclyde University - a une expérience étendue de l'utilisation des descriptions au niveau de la collection dans un environnement réparti dynamique, et des problèmes de terminologie qui y sont liés.

Il dispose également d'un riche environnement d'informations pour étudier la mise en place du projet pilote et son interaction avec les utilisateurs et les services. Cet environnement comprend : le catalogue partagé CAIRNS des universités, les bases de données de la NLS (BN d'Écosse), de la NGfL, de la SLAINTE et de la GDL (Glasgow Digital Library), un mécanisme d'aménagement par domaines des grandes collections, la base de données SCONE, le serveur électronique OAI (Open Archives Initiative), le New Opportunities Fund et d'autres projets de bases de données numériques, ainsi que la capacité à monter d'autres bases de données selon la norme Z39.50.

D'autres participants, en particulier UKOLN, mda, NCA, le RDN et la DNER, ainsi que les experts en terminologie HILT ajoutent encore de la profondeur et de l'envergure à l'équipe.

En outre, OCLC a accepté d'aider l'étude en fournissant un accès au mapping LCSH-DDC et à l'expertise qui y est liée. Le CDLR travaille aussi étroitement avec les dix établissements de formation continue de Glasgow membres du projet de bibliothèque numérique (GDL) du RSLP.

Réaliser TeRM

Pour les buts de ce projet, le prototype TeRM devrait être construit en utilisant le logiciel commercial Wordmap⁹, connu (par l'expérience de HILT phase I) pour fournir une bonne illustration initiale du genre d'équipements requis pour le prototype. Ceci n'implique pas qu'on préfère ce logiciel ou ce fournisseur, ni même qu'on préfère un produit commercial à un produit "maison" ou à une approche freeware. Le projet viserait à développer de complètes spécifications à travers des activités d'évaluation menées par des groupes appropriés d'utilisateurs et d'établissements, ainsi que par des experts externes. Il comparerait alors *tous* les produits appropriés disponibles, après avoir passé en revue les solutions commerciales courantes et autres. Wordmap serait parmi ceux capables d'offrir un logiciel qui saurait répondre à un nombre important de spécifications, mais ne serait pas favorisé. La question de savoir si une approche propre commune pourrait être préférable à l'achat d'un produit commercial serait également examinée.

Il y a de bonnes raisons qui justifient l'utilisation d'une partie spécifique d'un logiciel commercial à cette étape du développement. L'expérience de HILT phase I suggère que les participants au projet trouvent plus facile de discuter les conditions d'un tel service en utilisant un exemple réel sur lequel ils peuvent s'appuyer. Nous jugeons donc essentiel de réaliser un prototype illustratif très tôt dans le projet pour aider à susciter l'intérêt et l'attention des utilisateurs et autres parties et leur donner un environnement pratique concret pour envisager

⁴ See <http://cairns.lib.strath.ac.uk/> - Z39.50 catalogue including universities, NLS, NGfL, and others

⁵ See <http://scone.strath.ac.uk/> and <http://seed.cdlr.strath.ac.uk/>

⁶ See <http://scone.strath.ac.uk/service/index.cfm>

⁷ See <http://hilt.cdlr.strath.ac.uk/>

⁸ See <http://cdlr.strath.ac.uk/>

⁹ See www.wordmap.com

et considérer le problème. Wordmap sera employé parce que nous voulons avoir un outil de travail réel à une étape précoce pour juger de l'interaction avec les utilisateurs et les fournisseurs de services. Essayer d'établir toutes les conditions de réalisation avant de mettre en application un prototype dans un vrai contexte n'aurait pas été suffisamment stimulant pour les utilisateurs et les fournisseurs de services et n'aurait pas permis, pensons-nous, aux spécifications d'émerger pleinement. Il s'agit donc d'une approche pragmatique qui nous permettra d'évaluer les vraies utilisations et les problèmes d'une manière opportune, tout en évitant également la perte de capacités et le risque qu'implique un développement à partir de rien avant d'avoir pu établir les conditions optimales.

Terminologies et questions liées à la terminologie

La première illustration de TeRM serait basée sur les terminologies du RDN¹⁰, sur les terminologies disponibles en tant qu'éléments des taxonomies de Wordmap, qui incluent notamment un ensemble de termes utilisés par les internautes en général, ainsi que sur une sélection de sous-ensembles des LCSH, de la DDC, du thésaurus de l'Unesco et de l'AAT. OCLC fournira les équivalences LCSH-DDC et peut-être les équivalences entre la DDC et les vedettes-matière Conspectus¹¹. Le thésaurus de l'Unesco est disponible en ligne¹² et nous chercherons à obtenir des sélections de l'AAT à partir de sources manuelles. Le but serait d'avoir un mapping sélectif suffisant pour établir un premier prototype - et non un mapping terminologique exhaustif.

On prendrait aussi en considération les diverses questions soulevées par le rapporteur de HILT phase I, Leonard Will (HILT final Report, section 10), ainsi que deux questions supplémentaires :

1. La question de savoir si TeRM doit posséder une épine centrale

Un élément clé dans la fourniture d'un tel prototype sera de "traduire" la requête matière de l'utilisateur, telle qu'exprimée avec sa propre terminologie, dans les diverses terminologies effectivement employées dans l'environnement bibliographique, et de le faire d'une façon intelligente et utile. Cela demandera normalement un certain nombre d'interactions utilisateur-TeRM pour désambiguïser le ou les termes employés par l'utilisateur : (par exemple, s'il cherche "lotus", demande-t-il la fleur, la voiture, le logiciel, ou autre chose encore?)

La question qui se pose alors est de savoir si l'on a :

- a. une meilleure faisabilité
- b. une meilleure récupération par l'utilisateur (ce qui induit, notons-le, une obligation de rechercher à travers des barrières linguistiques)

si cette interaction a lieu entre l'utilisateur et un système matière central unique vers lequel tous les autres systèmes matière présents sont mappés dans TeRM,

ou bien si cette interaction a lieu entre l'utilisateur et chacun des systèmes matière tour à tour.

Si la meilleure réponse est dans le système central unique, reste encore la question de savoir lequel, s'il existe, atteindrait le mieux cet objectif, le plus susceptible d'être utilisé étant probablement la DDC (classification bien structurée et dont l'organisation hiérarchique est déjà traduite dans un nombre significatif des principales langues).

¹⁰ See, as an indication, the list created by Andy Powell at <http://www.rdn.ac.uk/cgi-bin/browse>

¹¹ Conspectus subject headings are used in the CAIRNS collection strengths database

¹² See <http://www.ulcc.ac.uk/unesco/thesaurus.htm>

2. La question de savoir si la meilleure solution à long terme au problème de l'accès matière dans un réseau partagé ne serait pas plutôt d'identifier et d'adopter un système unique commun admis comme apte à couvrir tous les besoins dans tous les domaines, ou bien encore d'envisager un système entièrement nouveau, ou un système existant mais modifié pour satisfaire les exigences d'un modèle préalablement accepté.

Il s'agit là, essentiellement, d'un prolongement de l'analyse coût-efficacité liée à l'idée d'un service de mapping terminologique - une évaluation extérieure de l'intérêt de TeRM pour l'ensemble de la communauté concernée par comparaison avec d'autres approches possibles.

Translated by: Max Naudi

Etablir l'environnement de recherche

L'environnement de recherche pourrait être réalisé par l'ajout d'une sélection de la DNER et d'autres collections, en incluant les collections du RDN, des archives, des musées, et une collection locale OAI, jusqu'à une copie de la base de données des collections SCONE¹³ pour constituer pour HILT phase II une base de données-test avec un environnement de recherche utilisant le mécanisme dynamique CAIRNS et diverses autres facilités. Le but serait d'utiliser les systèmes d'accès matière présents pour interroger les collections dans l'environnement documentaire, et d'utiliser le prototype TeRM pour désambigüiser les termes de l'utilisateur et résoudre les différences entre les systèmes matière. Une gamme de base des différents types d'utilisateurs serait utilisée, globalement associée à un noyau de sujets selon les centres d'intérêt, mais représentant un contexte varié d'utilisateurs à un niveau local, régional, national et britannique dans son ensemble (en général et selon les noyaux sujets)¹⁴. Le but serait de faire le lien entre TeRM et les mécanismes d'aménagement existants si c'est possible (l'expérience de CAIRNS suggère que oui) ou de simuler cet aspect si ça ne l'est pas (ce qui serait moins élégant mais suffisant pour les buts du projet).

Les engagements de HILT Phase II

Les engagements indiqués pour HILT phase II sont :

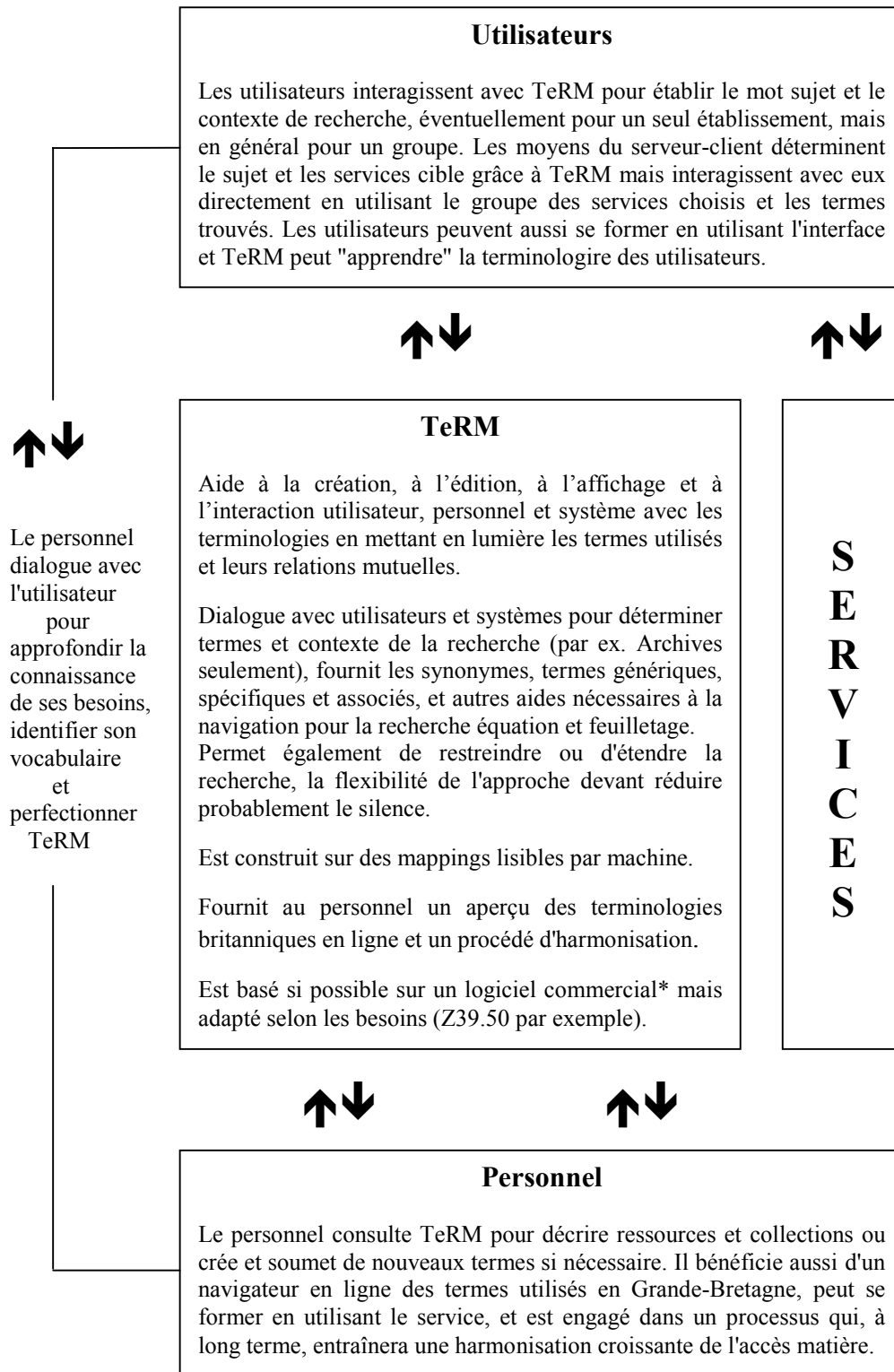
1. Une meilleure compréhension des problèmes et des besoins des utilisateurs de la formation continue et de l'enseignement supérieur en ce qui concerne la recherche matière dans le futur environnement documentaire (IE) du JISC, à la fois en direction des services internes au JISC et dans une perspective de diffusion au sein de la communauté dans son ensemble.
2. Une compréhension approfondie des exigences des équivalences terminologiques dans la DNER et les services britanniques associés, prenant en compte les considérations locales, régionales, nationales, internationales et "matière", ainsi que celles de la formation continue, de l'enseignement supérieur, des archives, bibliothèques, musées et autres services électroniques.
3. Un prototype opérationnel de service terminologique pour l'IE du JISC (avec des fonctionnalités limitées au départ et des possibilités de service complet exigeant probablement un changement de logiciel).
4. Les conditions requises, les coûts d'installation et de maintenance, la comparaison des avantages et des inconvénients du service envisagé, en fonction des exigences de fonctionnalité et de terminologie à la fois pour les utilisateurs et pour le travail de machine à machine.
5. Un rapport final sur le projet ainsi que des recommandations appropriées.

Si les subventions prévues sont versées (seul un avis informel a été donné au moment où ces lignes sont écrites), HILT phase II commencera en mai 2002.

¹³ <http://scone.strath.ac.uk/Service/Index.cfm>

¹⁴ Précisément : noyaux sujets pour une université, un collège de formation continue, l'enseignement supérieur, la formation continue à Glasgow, etc.

Annexe A. Diagramme interactif TeRM (Terminologies Route Map)



*Note: Pour des exemples, voir : www.wordmap.com avec www.oingo.com et vivisimo.com