

La réutilisation des données de thésaurus de bibliothèques comme ontologies pour le secteur public.

[French] translation of the original paper: “Reuse of library thesaurus data as ontologies for the public sector”.

Translated by: Agnès Simon, Bibliothèque nationale de France, Paris, France.

Le texte de ce document a été traduit en français et des différences avec le texte initial peuvent apparaître. Cette traduction est fournie dans un objectif uniquement de consultation.

Mikko Lappalainen

Bibliothèque de recherche, Bibliothèque nationale de Finlande, Helsinki, Finlande.
mikko.lappalainen@helsinki.fi

Matias Frosterus

Services de réseau de la bibliothèque, Bibliothèque nationale de Finlande, Helsinki, Finlande.
matias.frosterus@helsinki.fi

Susanna Nykyri

Bibliothèque de recherche, Bibliothèque nationale de Finlande, Helsinki, Finlande.
susanna.nykyri@helsinki.fi



This is a French translation of “*Reuse of library thesaurus data as ontologies for the public sector*” copyright © 2014 by Agnès Simon. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>

Résumé :

Au Printemps 2013, la Bibliothèque nationale de Finlande a pris en charge la construction d'une infrastructure d'ontologie de niveau national pour tout le secteur public. Le service d'ontologie en open source Finto et l'ontologie générale finlandaise YSO sont au centre de cet effort. Cette ontologie repose sur le Thésaurus général finlandais, à l'origine développé dans les années 1980 à la Bibliothèque nationale, et portant essentiellement sur l'indexation des livres. Le passage de termes d'indexation utilisés dans le thésaurus, aux concepts ontologiques YSO, a rendu possible l'alignement de cette ontologie avec d'autres ressources. Des transformations similaires ont été utilisées pour des thésauri d'autres domaines. En liant les différents thésauri traduits sous formes d'ontologies, il est devenu possible d'harmoniser des métadonnées entre différentes bases de données. YSO et Finto sont actuellement utilisées dans des projets pilotes dans plusieurs organisations. Aujourd'hui la Bibliothèque nationale de Finlande acquiert un nouveau rôle de pilotage du développement et des futures pratiques d'indexation, dans le monde du web de données.

Mots-clés: Thésauri, Ontologies, Métadonnées, Gestion de contenu, Web de données

1 INTRODUCTION

L'omniprésence d'Internet et l'importance qu'il donne à la recherche en texte libre ont un temps porté à croire que le rôle des annotations allait diminuer. Il est toutefois apparu assez rapidement, qu'en raison de l'accroissement exponentiel de la masse de données, Le besoin d'annotations n'a jamais été aussi criant, mais dans un nouveau format, lisible par les machines, qui puisse aider les utilisateurs humains à comprendre les données grâce à une meilleure automatisation. C'est un, appel auquel les bibliothèques, avec leur grande expertise de curateurs de connaissances, sont les mieux placés pour répondre.

La construction d'une infrastructure web sémantique pour la Finlande a été originellement initiée par le projet de recherche FinnONTO¹ (2003-2012) de l'Université Aalto et l'Université d'Helsinki. Le projet a réalisé plusieurs portails sémantiques, un service d'ontologie, un certain nombre d'outils pour utiliser les ontologies dans des annotations, et a aussi transformé en ontologies plusieurs thesauri clés utilisés dans différents domaines.

En se fondant sur les résultats du projet FinnONTO, la Bibliothèque nationale de Finlande a commencé son projet actuel ONKI² en collaboration avec le Ministère des Finances et le Ministère de l'Éducation et de la Culture en 2013, avec pour objectif de construire un service d'ontologie à l'échelle nationale, assorti d'un jeu de vocabulaires et d'ontologies répondant à divers besoins d'indexation et d'interconnexion. Service utilisé pour la publication et l'utilisation d'ontologies comme de thesauri, FINTO³ a été lancé en janvier 2013. Le projet développe en outre l'ontologie finlandaise générale YSO⁴, qui a pour objectif de fournir un ensemble global de concepts communs à tous les domaines.

En comparaison de l'axe de recherche du projet FinnONTO, le projet ONKI accorde plus d'importance à la fiabilité, la facilité d'utilisation et le point de vue d'une vraie personne cherchant de l'information. En outre, le principe qui sous-tend les ontologies a changé : passant de la tentative de refléter précisément le monde tel qu'il est, à la mise à disposition d'un outil pour l'annotation et la recherche d'information, aussi utile que possible. Le changement de paradigme était nécessaire, dans la mesure où la stabilité et la facilité d'utilisation sont devenues des enjeux prioritaires du service, pour en faire « un socle de construction » fiable pour d'autres systèmes. L'attention prêtée à la recherche d'information résulte d'une meilleure compréhension des besoins des vraies applications pour les utilisateurs finaux.

Avec la migration des thesauri vers les ontologies et des termes vers les concepts et des termes en concept s'ouvrent de nouvelles possibilités d'interopérabilité et de réutilisation. Les termes portent différentes significations dans différents domaines, mais en explorant plus avant les concepts réels derrière les termes, en donnant aux concepts des identifiants uniques et en définissant des relations avec d'autres concepts, il devient possible d'aligner Les divers vocabulaires issus de domaines différents. Un total de quinze thesauri, en vigueur dans différents domaines, ont été transformés en ontologies et alignés dans un nuage harmonisé d'ontologies interconnectées. Actuellement le nuage d'ontologies interconnectées et le service d'ontologie Finto sont en utilisation pilote dans diverses organisations du secteur public.

Pour transformer des thesauri en ontologies, notre approche insiste sur l'importance de maintenir un lien clair vers les annotations faites avec les vocabulaires préexistants. Un autre aspect essentiel est la répartition de la production de l'ontologie entre des organisations expertes, en charge des thesauri d'origine. Cependant, plutôt que de se contenter d'aligner plusieurs ontologies conçues indépendamment, nous souhaitons encourager des processus de coopération et une plus forte interconnexion, garants de la qualité des hiérarchies ainsi obtenues.

Nous évoquerons dans un premier temps la transformation du Thesaurus général finlandais, YSA, largement utilisé, en Ontologie générale finlandaise, YSO. Nous décrirons aussi comment les termes ont été explicités en concepts précis, grâce à des relations exploitables par les machines et une hiérarchie complète, et en quoi cela a été une étape indispensable pour l'interconnexion des données entre différentes organisations. Dans la troisième section, nous donnerons un panorama du service d'ontologie national, qui a été développé dans le projet ONKI, et nous détaillerons le nuage d'ontologies liées. Enfin, dans la quatrième section, nous étudierons le rôle nouveau qui s'est ouvert aux bibliothèques nationales à l'ère des données liées (*linked data*).

¹ Voir <http://www.seco.tkk.fi/projects/finnonto/>

² Voir <https://wiki.helsinki.fi/display/ONKI/>

³ Voir <http://finto.fi/en/>

⁴ Voir <http://finto.fi/ysa/en/>

2 DE TERMES VAGUES AUX CONCEPTS LIÉS

Au cœur du projet ONKI et du service Finto se trouve l'Ontologie générale finlandaise YSO, elle-même reposant sur le Thésaurus général finlandais YSA, à l'origine développé dans les années 1980 à la Bibliothèque nationale de Finlande, et principalement conçu pour l'indexation par chaînes de caractères des documents imprimés de bibliothèques.

YSO fut initialement construit élaboré dans le cadre du projet de recherche FinnONTO, par le biais d'une transformation des termes d'indexation d'YSA en concepts, grâce à la définition des unités de pensée qui les sous-tendent. Des identifiants uniques (URIs) ont été attribués aux nouveaux concepts, pour les identifier indépendamment des termes de langage spécifiques. Afin de rendre l'ontologie bilingue, le Thésaurus général suédois Allärs⁵ a été utilisé pour fournir les formes suédoises des concepts. Des formes anglaises furent également données aux nouveaux concepts, par la traduction des termes finnois en anglais.

Une nouvelle ontologie générique a été réalisée sur la base de modèles existants comme DOLCE⁶ et BFO⁷ et tous les concepts ont été placés dans une seule structure hiérarchique arborescente, en utilisant des relations standardisées entre les concepts. L'ontologie qui en résulte (YSO) a été présentée sous une forme lisible par les machines, de telle sorte qu'elle pouvait être facilement utilisée dans diverses applications, conformément aux principes du web de données.

Cependant, l'« ontologisation » initiale d'YSA a été poursuivi en mettant l'accent sur son aspect recherche, au détriment d'une réelle facilité d'utilisation de l'ontologie finale. Ainsi, les premières versions d'YSO ont-elles critiquées par le monde des bibliothèques, qui les jugeait obscures et inutilisables, et le remplacement d'YSA par YSO fut considéré comme peu probable. Les utilisateurs se plaignaient du niveau générique d'YSO, incohérent et source de confusion, ainsi que des nouveaux concepts, créés pour compléter la hiérarchie, qui étaient mal choisis. Il en résulta que le caractère globalement intuitif de l'ontologie fut perdu. Il s'est révélé qu'en rendant YSA lisible par les machines, on l'avait rendu lisible par les machines seulement.

Dans le cadre du projet ONKI, on a porté une attention particulière à rendre YSO intuitif et facile d'utilisation, sans perdre le bénéfice des données sémantiques lisibles par les machines. Ci-dessous, nous abordons des principaux défis rencontrés dans la poursuite de cet objectif.

2.1 Transformer un thésaurus en ontologie : les défis clés liés au contenu

YSA est un thésaurus généraliste en finnois dont le contenu couvre tous les domaines à un niveau générique. Son objectif premier est d'être utilisé comme outil d'indexation dans les bibliothèques, les archives et les musées de Finlande, bien qu'en pratique, son usage en dehors des bibliothèques ait été limité. Le thésaurus est maintenu par la Bibliothèque nationale de Finlande. YSA a été utilisé pour indexer des publications finlandaises depuis 1987. Il a également une traduction suédoise, Allärs, utilisée depuis 1996.

Dans un premier temps, YSA a été développé et maintenu dans l'optique des pratiques et des besoins traditionnels des bibliothèques et n'est donc pas indiqué, en tant que tel, pour servir aux utilisateurs plus hétérogènes du web sémantique⁸ ou répondre à leurs besoins variés. Le Web sémantique inclut différents types de données, de publications, d'acteurs et de discours, qui devraient tous être reliés entre eux afin d'être exploitables dans différentes bases de données et organisations.

⁵ Allärs était déjà aligné avec YSA.

⁶ Masolo et al. 2003.

⁷ Voir <http://www.w3c.org/Designissues/LinkedData.html>

⁸ Selon le W3C, « Le web sémantique porte sur deux points. Il s'agit de formats communs pour l'intégration et la combinaison de données de sources diverses, là où le web traditionnel se concentre principalement sur l'échange de documents. Il s'agit aussi d'un langage pour enregistrer la façon dont les données sont reliées avec les objets du monde réel. Cela permet à une personne ou à une machine de partir d'une seule base de données, puis de traverser un nombre sans fin de bases de données, qui sont reliées non par des fils mais parce qu'ils concernent le même objet. » <http://www.w3c.org/2001/sw/>

En pratique, le transfert d'un thésaurus vers une ontologie implique deux changements majeurs : premièrement, l'attention se focalise au-delà du niveau du terme pour se situer au niveau du concept (« unité de pensée »⁹) et, deuxièmement, la structure hiérarchique est complète et cohérente. Dans les thesauri, les termes ne représentent pas nécessairement des concepts isolés et bien définis, mais peuvent contenir des ambiguïtés sous-jacentes. En outre, les relations entre les termes d'un thésaurus n'explicitent pas nécessairement la différence entre des relations, par exemple de type « est-un » et « est-une-partie-de ». Un thésaurus (comme YSA) rend possibles des pratiques imprécises, si la structure hiérarchique n'est pas soigneusement définie et complète. L'utilisateur lit un thésaurus d'un certain point de vue et ne pense pas forcément aux autres manières de comprendre le terme, ce que la structure d'une ontologie permet.

L'imprécision est devenue un problème fréquent au niveau du corpus, précisément parce qu'YSA est un thésaurus général. Par exemple, la forme retenue « aura » n'a dans YSA ni terme plus large, ni terme plus précis, mais comme groupe thématique elle correspond à la fois à la classe 48 pour « médecine »¹⁰ et 50 pour « folklore »¹¹. Si les utilisateurs perçoivent les groupes thématiques et connaissent bien les principes et les pratiques d'indexation, il peut leur être évident que le terme dispose de deux sens différents et qu'il représente, par conséquent, deux concepts effectivement différents. S'ils sont familiers de ces pratiques, la multiplicité des résultats de recherche ne les surprendra pas, et/ou ils savent affiner les résultats en enrichissant leur stratégie de recherche par des concepts pertinents¹².

Dans l'ontologie YSO, le terme du thésaurus « aura » a été divisé en deux concepts et, par conséquent, en deux entrées différentes : « aura (phénomène paranormal) (fr) » et « aura (symptômes) (fr) » (faisant référence à un trouble de la perception typiquement lié à la migraine). (Voir ci-dessous figure 1). Ainsi, ces termes sont liés à des concepts génériques différents, et comprennent tous deux des notes précisant leur signification dans l'ontologie. Ce modèle « un concept, un terme » permet aux personnes qui indexent de faire une différence conceptuelle pendant le processus d'indexation, évitant ainsi à l'utilisateur des résultats de recherche flous ou inutilement nombreux.

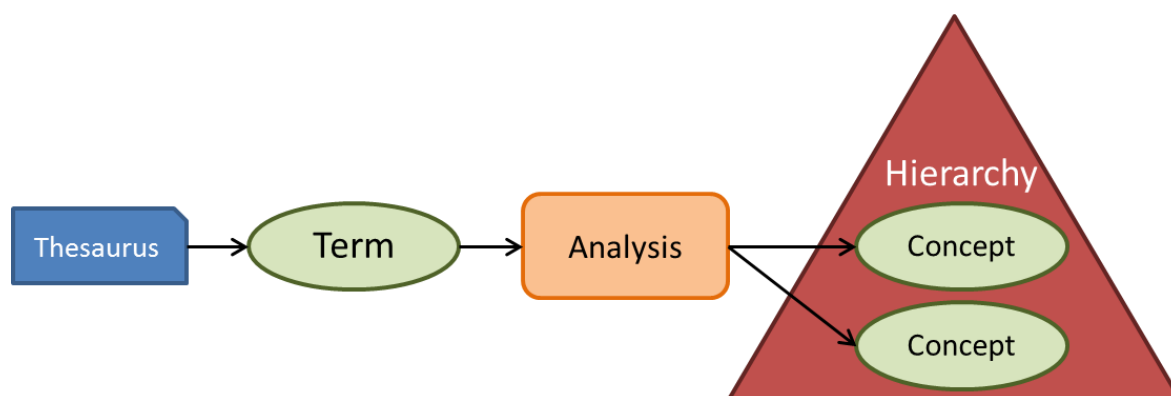


Figure 1 : Le processus d' « ontologisation »

⁹ Selon ISO (ISO 5963-1985,1) un concept est une unité de pensée et « Le contenu sémantique d'un concept peut être exprimé à nouveau par une combinaison d'autres concepts différents, qui peut varier d'une langue ou d'une culture à une autre ». (Ibid.) ISO -standard 25964-1 :2011 (p.3) précise que les concepts existent dans l'esprit comme des entités abstraites et indépendantes des termes utilisés pour les exprimer.

¹⁰ [En Finnois : Lääketiede. Anatomia. Fysiologia. Patologia. Tautioppi. Psykiatria. Hammaslääketiede. Eläinlääketiede. Farmasia. Kauneudenhoito. Hoitomenetelmät]

¹¹ [En Finnois : Kansanperinne. Perinnetieteet. Kulttuuriantropologia. Okkultismi]

¹² Melinda, le catalogue collectif des bibliothèques finlandaises avait quatre différentes significations de « aura » (représentant ainsi quatre différents concepts) dans les cinq premières notices obtenues avec la recherche sujet « aura ».

Les ontologies, comme les thésaurus, sont des vocabulaires contrôlés. Les normes ISO 5964-1985 et plus récemment ISO 25964-1 :2011 représentent deux des moyens principaux pour exercer le contrôle de vocabulaire dans un thésaurus : a) Les champs des concepts et termes sont délibérément restreints à des significations choisies, et b) quand le même concept peut être exprimé par deux synonymes ou quasi-synonymes ou plusieurs dans une même langue, un de ces termes est généralement choisi comme forme retenue.¹³

Bien que le but du projet FinnONTO ait été de fonctionner au niveau du concept, il n'a pas été complètement atteint. Premièrement, plusieurs termes dans le corpus YSA avaient des significations vagues, et deuxièmement, YSO avait été construit dans une optique plus terminologique que fonctionnelle. Ceci a donné lieu à des solutions qui ne renaient pas les éléments sémantiques centraux du concept, afin de se conformer à l'utilisation effective d'YSO, c'est-à-dire un outil d'indexation et de recherche d'information.

Dans la version FinnONTO d'YSO, certains termes ne reposaient pas sur un usage avéré (attesté par la documentation), mais ambitionnaient de recouvrir tous les sens possibles associés à un terme, de façon exhaustive. Beaucoup de concepts étaient divisés en plusieurs concepts, à cause de cette pratique, et par conséquent, YSO incluait aussi des termes qui ne renvoyaient à aucun résultat pour des recherches habituelles dans la base de données (c'est-à-dire que leur existence dans YSO n'est pas fondée sur des besoins d'indexation réels). Dans l'actuel projet ONKI, le même corpus a été considéré davantage du point de vue de sa fonction – servir d'outil de récupération et d'indexation – faisant ainsi un pas en arrière (vers son modèle de thésaurus). En pratique, cela signifie que les termes polysémiques n'ont pas automatiquement des points d'accès distincts pour chaque signification, mais seulement pour celles dont on a besoin pour l'indexation. Cela signifie aussi que l'analyse du concept s'attache plus à l'utilisation d'YSO en tant qu'outil pour indexer et trouver l'information. La fonction de l'ontologie détermine le point de vue qui préside à l'analyse du contenu, en déterminant en dernier recours la représentation du concept dans la hiérarchie.

Le passage des thesauri aux ontologies implique également de se tourner vers une acception des concepts qui s'affranchisse de la langue. Son corpus, le Thésaurus général finlandais YSA, est en finnois, mais il a aussi une version en suédois, Allårs, qui a été construite comme un vocabulaire séparé¹⁴. La transformation en ontologie a facilité l'ajout d'équivalents dans différentes langues, puisqu'on se concentre désormais clairement sur des concepts et non plus sur des termes. Dans le développement du corpus, le finnois et le suédois ont un statut équivalent, mais l'anglais possède un statut de langue de traduction. Les traductions en anglais ont permis à YSO de s'aligner avec les vedettes de la Bibliothèque du Congrès (LCSH). Dans un premier temps, des libellés anglais (c'est-à-dire, des termes) furent attribués aux concepts finnois et, dans un second temps, on trouva des équivalents de la LCSH.

Si les mots et les significations sont traduisibles, ils peuvent toujours être vecteurs de différentes valeurs ou visions du monde dans des cultures différentes, et ainsi les sens des mots sont ouverts à interprétations¹⁵. Bien que multilingue, YSO reste mono-culturel, associé au contexte et au point de vue finlandais, reflet de sa société et de sa culture (par exemple, le système éducatif, l'idée de ce que l'on entend par « famille », etc.).

¹³ « Le champs des termes est délibérément restreint à des significations choisies. À la différence des termes d'un dictionnaire, qui peut être accompagné d'un certain nombre de définitions différentes reflétant l'usage commun, chaque terme d'un thésaurus est généralement restreint à la seule signification qui sert le mieux les besoins d'un système d'indexation. La structure d'un thésaurus, en particulier la diffusion des relations hiérarchiques, indiquent fréquemment le sens prévu d'un terme. Si la technique n'est pas suffisamment explicite, une définition ou un commentaire devrait être ajoutée au terme. Ceci doit établir le sens choisi et peut aussi indiquer d'autres sens, qui sont reconnus dans le langage naturel, mais qui ont été explicitement exclus pour des raisons d'indexation ; quand le même concept peut être exprimé par deux synonymes ou plus, un de ces termes est en général utilisé comme terme préféré (---) qui est alors utilisé pour l'indexation, de manière cohérente. On doit faire référence au terme préféré pour tout synonyme, lequel pourrait servir de point d'accès à un usager. » (ISO 5964,1985, 5)

¹⁴ La Finlande a deux langues nationales, le finnois et le suédois. Le finnois est la langue maternelle de 90%, et le suédois, de 5 % de la population finlandaise.

¹⁵ Katan 2004, voir plus sur Nykyri 2010.

De plus, un des défis centraux de l' « ontologisation » est le temps. Alors que de nouvelles possibilités émergent, il a également fallu se soucier des anciennes pratiques. Dans la mesure où YSA a une longue histoire d'indexation, une attention particulière a dû être accordée à l'indexation faite avec le thésaurus YSA – il est absolument vital que les anciennes ressources restent accessibles et ne soient pas perdues du fait du nouvel outil d'indexation, YSO.

2.2 Choix techniques et détails structurels

Un des principes clés du web de données est la lisibilité par les machines. En substance, cela signifie que les propriétés liant les différentes ressources ont des sens explicites, définissant nouvelles relations un moteur d'inférence est en droit de déduire à partir de relations existantes. Initialement les ontologies étaient décrites en OWL est devenu SKOS par la suite.¹⁶ Cependant des relations sémantiques de type strict « est-un » ont été maintenues pour les relations hiérarchiques, même si une propriété plus souple skos :broader a été employée pour désigner les hiérarchies. Cela signifie que la hiérarchie en YSO peut être utilisée transitivement comme une sous-classe hiérarchique, de manière fiable.

La transitivité requiert d'avoir une ontologie supérieure bien pensée et fonctionnelle. Une des premières tâches du projet ONKI a été de revoir le niveau générique d'YSO pour qu'il fonctionne comme prévu. Prenant en compte les critiques faites à YSO par le passé, on jugea important que l'ontologie générique, tout en étant lue et interprétée par des machines, pût également l'être par des humains de façon intuitive. Puisque YSO fonctionne comme une ontologie générique pour les autres ontologies de domaines spécifiques (voir section 3.2.), les classes génériques ont dû être développées afin d'être utilisables dans différents domaines de connaissance.

Une structure hiérarchique complète garantit aussi un plus grand nombre de requêtes automatiques et Permet d'améliorer la précision et la mémoire des requêtes grâce à des méthodes¹⁷ sophistiquées. Dans la mesure où le projet ONKI a été centré sur la construction d'ontologies qui peuvent être réellement utilisées pour trouver l'information les choix hiérarchiques d'YSO ont été considérés comme cruciaux.

Au début du projet ONKI, un grand nombre d'ontologie d'ontologies génériques disponibles ont été passées au crible, et des catégories métaphysiques ont été étudiées, afin d'évaluer les différences de choix effectuées par les différents modèles. Il a été décidé qu'un nouveau modèle générique devait satisfaire trois critères : être cohérent de part en part pour que les déductions automatiques puissent produire des résultats utiles ; se fonder sur des modèles utilisés et approuvés pour simplifier l'alignement entre YSO et d'autres ressources ; et enfin, définir les concepts génériques de telle sorte qu'ils ne soient pas seulement lisibles par les machines, mais aussi compréhensibles par les humains.

Le nouveau modèle générique d'YSO repose sur l'Ontologie descriptive pour la linguistique et l'ingénierie des connaissances (*Descriptive Ontology for Linguistic and Cognitive Engineering*, DOLCE)¹⁸, principalement parce que son but n'est pas de « modéliser le monde » mais de clarifier des hypothèses ontologiques cachées dans des ressources lexicales existantes. Le fait que DOLCE soit déjà aligné avec certaines grandes ressources lexicales comme WordNet a aussi été considéré comme un plus. Un autre point à prendre en considération est l'utilisation d'un identifiant pérenne unique établir la distinction entre concepts (annotations) et étiquettes (termes).

YSO utilise des URIs actionnables, qui affichent la page du concept pertinent dans Finto, quand on y accède par un navigateur, et les données en RDF, quand une application tierce utilise les interfaces machine de Finto. La structure choisie pour les URIs ne contient pas d'information lisible par un humain, afin qu'ils conservent leur indépendance linguistique et pour éviter toute pression pour modifier un identifiant en fonction d'un changement de la forme préférée d'un concept. Par exemple, l'URI du concept « chat » dans YSO est <http://www.yso.fi/onto/yso/p193>. Il ne montre donc pas de préférence pour une langue donnée, ni ne subit aucune pression pour changer, même si la forme préférée du concept de chat était amenée à changer.

¹⁶ Voir <http://www.w3c.org/2001/sw/wiki/OWL> et <http://www.w3c.org/2004/02/skos>

¹⁷ Voir par exemple Agraval et al. 2009

¹⁸ Masolo et al., 13-41

3 PERMETTRE LES REUTILISATIONS

Dans la partie précédente, nous avons évoqué en détail la transformation du thesaurus général en une ontologie générale. Cependant, afin de rendre utilisable cette ontologie en dehors du périmètre de la Bibliothèque nationale, deux choses étaient nécessaires. Premièrement, un service permettant la publication et l'utilisation d'ontologies d'une part, service stable à l'interface moderne pour faciliter autant que faire se peut son intégration dans différentes applications ; l'interconnexion et l'harmonisation de thesauri utilisés par d'autres organisations avec l'ontologie générique d'autre part. Le tout résultant en un nuage d'ontologies reliées définissant les relations entre les concepts utilisés dans des annotations et permettant ainsi de croiser les recherches entre plusieurs bases de données.

3.1 Le thesaurus Finto et le service d'Ontologie

Pour commencer, le projet FinnONTO a réalisé un service d'ontologie¹⁹, à savoir un répertoire centralisé d'ontologies qui soit accessible par un navigateur ou différentes interfaces, permettant l'intégration dans d'autres systèmes²⁰. Quand SKOS a été développé, il a été adopté pour la publication d'ontologies légères et en 2012, une nouvelle version simplifiée du service d'ontologie fut entièrement reconstruite, en s'appuyant sur SPARQL²¹. Quand la Bibliothèque nationale de Finlande lança le projet ONKI en 2013, cette dernière incarnation fut choisie comme base de la nouvelle version en production du service.

Afin de se distinguer clairement des précédentes versions, qui étaient destinées à la recherche par des utilisateurs, le nouveau service fut nommé Finto, service de thesaurus et d'ontologie finlandais (Finnish Thesaurus and Ontology service). Le logiciel libre sous-jacent, Skosmos, est en cours de développement dans GitHub²². Finto est une implémentation spécifique de Skosmos – un outil de recherche de vocabulaire et de thesaurus utilisant SKOS et SPARQL – mais il peut être utilisé pour installer un service d'ontologie n'importe où.

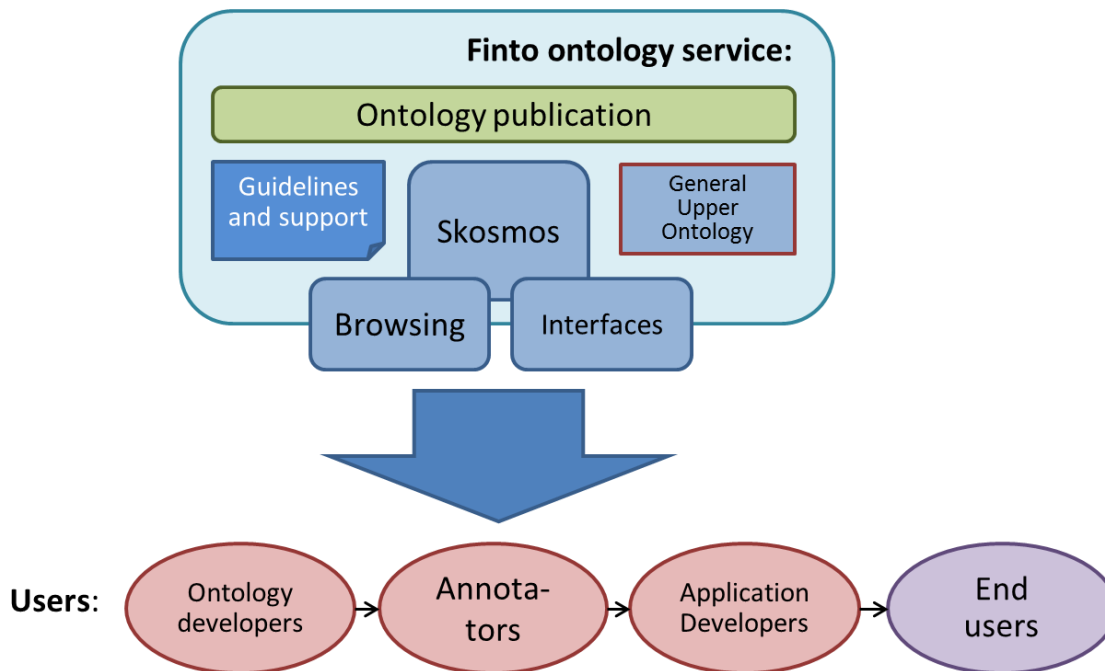


Figure 2 : Le service d'ontologie Finto et ses groupes d'utilisateurs

Les composants du service Finto sont représentés dans la **figure 2**. La publication d'ontologies en est la fonction majeure : il s'agit de mettre à disposition un lieu centralisé, où l'on puisse trouver et utiliser aussi bien les

¹⁹ Voir <http://onki.fi/>

²⁰ Viljanen et al. 2009

²¹ Suominen et al. 2012

²² Voir <https://github.com/natlibfi/skosmos>

ontologies que les méthodes et interfaces communes nécessaires à leur accès et réutilisation. Ceci rend l'intégration de différentes ontologies dans le même système aussi facile que possible, permettant par exemple, de créer un système d'annotation où différents champs sont liés à différentes ontologies.

En plus d'une plateforme technique, Finto offre par ailleurs des guides généraux de bonnes pratiques et des aides pour transformer des thesauri en ontologies légères SKOS, pour les maintenir ou les mettre à jour. La Bibliothèque nationale de Finlande a également publié l'ontologie générale finlandaise YSO, qui contient les concepts annotés requis dans plusieurs domaines.

La **figure 1** inclut également les groupes d'utilisateurs de Finto. En premier viennent les producteurs d'ontologies qui publient des ontologies et peuvent utiliser l'interface de navigation pour examiner d'autres ontologies, tout en travaillant sur les leurs. En deuxième viennent les annotateurs qui intègrent Finto dans leur environnement d'annotation, mais qui ont aussi la possibilité de parcourir au besoin les hiérarchies ontologiques. Le troisième groupe d'utilisateurs comprend les développeurs d'application souhaitant intégrer des ontologies et des annotations ontologiques dans leurs applications. Enfin, les utilisateurs finaux n'utiliseront probablement pas Finto directement, mais plutôt au travers d'applications fabriquées par le troisième groupe d'utilisateurs. Notons que la Bibliothèque nationale de Finlande prend en compte les usagers de tous ces groupes, Finto étant à l'essai en interne.

3.2 KOKO, le nuage d'ontologie liée

Dans notre deuxième partie, nous avons décrit la transformation du thesaurus YSA en ontologie YSO. Suivant les mêmes principes, le projet FINNONTO a aussi transformé quinze thesauri spécialisés en vigueur dans différents domaines en ontologies de domaines.

La transformation s'est opérée conjointement avec les experts des domaines qui développent les thesauri sous-jacents et les ontologies qui en résultaient ont été converties en YSO. Ainsi YSO fut utilisé comme une ontologie générique de haut niveau offrant une hiérarchie partagée, laquelle a été ensuite étendue par chaque ontologie de domaine. Lorsque toutes ces ontologies sont combinées, le résultat final est un nuage d'ontologies liées appelé KOKO.²³

La quantité variable de ressources disponibles a représenté un défi crucial pour les institutions chargées de maintenir des nouvelles ontologies de domaine. Dans la mesure où l'expertise de domaine nécessaire pour maintenir les ontologies de domaine est élevée, la coopération des organisations responsables du thesaurus original était d'une importance capitale. D'où, malgré tout, une infinité de différences dans les cycles de mise à jour des ontologies.

Si toutes les ontologies de domaine avaient été entre elles, tout changement dans l'une était susceptible d'affecter les quatorze autres. Il s'agissait d'une proposition intenable au vu des disparités de ressources. La solution retenue fut d'utiliser YSO comme un pivot central ou un ciment reliant entre elles toutes les ontologies de domaine tout en limitant le nombre de liens directs entre elles. De cette façon, les producteurs d'ontologies de domaine pouvaient ne s'attacher qu'aux modifications apportées à la seule ontologie YSO.

Le travail sur ce nuage harmonisé d'ontologies liées fut entamé pendant FinnONTO et est aujourd'hui développé dans le projet ONKI. L'objectif final est d'utiliser KOKO comme lien entre les annotations dans les jeux de données des différentes organisations, pour faciliter l'interopérabilité et briser les silos.

²³ Frosterus et al. 2013

Ontologie	Domaine	Nombre de concepts
YSO	Ontologie générique	24,800
MAO	Artefacts de musées	6,800
MUSO	Musique	1,000
TAO	Design	3,000
TERO	Santé	6,500
VALO	Photographie	2,000
AFO	Agriculture	7,000
JUHO	Gouvernement	6,300
KAUNO	Littérature	5,000
KTO	Linguistique	900
KITO	Recherche en littérature	850
KULO	Recherche culturelle	1,500
LIITO	Economie	3,000
MERO	Navigation	1,300
PUHO	Domaine militaire	2,000
TSR	Emploi	5,100

Table 1 : les ontologies composant KOKO

Les différentes ontologies comprises dans KOKO sont montrées dans la Table 1. Comme on peut le voir, elles couvrent un large panel de domaines variés et ont été construites à partir de thesauri en vigueur, de sorte que chacune possède, via les liens vers les thesauri, une grande quantité de données préalablement utilisées pour des annotations. La version actuelle de KOKO est une ontologie globale et harmonisée de 45 000 concepts alignés dans une seule hiérarchie.

KOKO est à l'essai comme vocabulaire d'annotation, notamment dans différents musées et à la Compagnie Nationale de Diffusion, c'est à dire dans des organisations susceptibles de gérer des ressources issues de tous les domaines possibles et imaginables.

4 UN NOUVEAU ROLE POUR UNE BIBLIOTHEQUE NATIONALE : PLUS DE CONTROLE, MIEUX REPARTI

Quand le projet ONKI a commencé au début de l'année 2013, il était évident que le travail initié dans le cadre d'un projet de recherche technique pouvait se révéler réellement utile pour une bibliothèque nationale, en dépit de la forte opposition du secteur des bibliothèques, qui avait chahuté le projet par le passé. Malgré son incomplétude, il s'agissait d'une opportunité à ne pas manquer.

L'adoption du projet impliquait d'endosser un rôle nouveau dans le contrôle de l'indexation à l'échelle nationale. Le monde avait énormément changé au cours des trois décennies écoulées depuis les débuts d'YSA à la Bibliothèque nationale. L'indexation, de manière générale était devenue accessible à tous : les gens annotent leurs propres articles de blogs, attribuent des hashtags à leurs actualités Twitter et annotent leurs photos pour les réseaux

sociaux. Le projet ONKI avec son penchant pour le web de se fait en quelque sorte le héraut de ce monde, et en même temps que des nouvelles possibilités de nouvelles responsabilités surgissent Notre nouveau rôle se déploie sur la scène bien plus large des données liées et des autorités (le terme me semble bienvenu) en général, bien au-delà des frontières du monde des bibliothèques.²⁴

En un sens, il était naturel qu'une bibliothèque, en particulier une bibliothèque nationale, prît la tête d'un projet qui allait construire un service et des outils couvrant les besoins d'indexation de tout le secteur public. Après tout, c'étaient bien les bibliothèques qui avaient derrière elles les décennies d'expérience en termes de construction de thesauri de de vocabulaires contrôlés et de leur utilisation en vue de l'indexation. Le défi majeur était de trouver une façon de combiner le savoir-faire de l'indexation traditionnelle de livre avec les meilleures idées et pratiques du web de données, et ainsi acquérir un rôle à la fois nouveau et traditionnel.

Bien qu'YSA ait été utilisé en-dehors de la Bibliothèque nationale, il s'agissait cependant toujours d'un système clos, au sens où il y n'y avait aucune réflexion sur la façon dont YSA était utilisée en dehors du secteur public et dont elle devait être interopérable avec les autres ressources d'indexation.

Le nouveau réseau comprenait de nouvelles catégories de joueurs, qui voulaient utiliser des ontologies toutes faites et disponibles via Finto, mais aussi lier leurs propres ressources d'indexation à d'autres, via le nuage d'ontologie KOKO. Ces nouveaux joueurs incluent des musées, des archives, des médias et des agences gouvernementales, avec lesquels, pour certains, la Bibliothèque nationale n'avait que peu collaboré par le passé. Afin d'orienter cette nouvelle forme de collaboration avec des partenaires, il a été décidé que la Bibliothèque nationale monterait un groupe d'intérêt spécial pour orienter le travail d'indexation de ce large réseau.

Au printemps 2014 la formation de ce groupe était toujours en cours et les arrangements finaux comme les procédures restaient à définir. Les premiers principes en ont cependant été explicités : l'idée est qu'à l'avenir, la Bibliothèque nationale fournisse le service (Finto) d'accès et d'utilisation des ressources d'indexation, et qu'elle constitue les principes de construction et de maintenance de toutes les ontologies reliées. L'objectif était que les organisations d'experts spécialisés soient toujours responsables des contenus des vocabulaires de domaines spécifiques.

À travers ce projet, la Bibliothèque nationale de Finlande a, d'une certaine façon, acquis un plus grand contrôle de l'indexation sur le plan national. Cela a été permis par l'adoption des principes du web de données, qui repose sur l'idée de personnes et d'organisations travaillant en commun pour le bénéfice de chacun. Bien que la Bibliothèque nationale fût l'acteur qui a construit le service d'utilisation commune des ontologies et a institué les principes fondateurs des ontologies nationales, il était clair que dans la forme finale de cette nouvelle infrastructure, la Bibliothèque nationale de Finlande ne serait qu'une entité parmi d'autres. En substance, les aspects organisationnels du projet ONKI s'attachaient à réformer et mettre en réseau : d'abord rassembler les parties en présence, puis les faire travailler ensemble sur un pied d'égalité.

L'expertise reconnue des bibliothèques dans la gestion de contenu et l'organisation de la connaissance offre une base naturelle pour l'interopérabilité entre domaines et l'harmonisation des métadonnées, que le projet ONKI avait pour objectif de développer. *In fine*, cela devrait conduire à une recherche d'information plus précise et plus complète par l'utilisateur.

5 REFERENCES

Agrawal, Rakesh; Gollapudi, Sreenivas; Halverson, Alan; Jeong, Samuel 2009: "Diversifying Search Results". <http://research.microsoft.com/apps/pubs/default.aspx?id=73931>

Baker, Thomas; Bermès, Emmanuelle; Coyle, Karen; Dunsire, Gordon; Isaac, Antoine; Murray, Peter; Panzer, Michael; Schneider, Jodi; Singer, Ross; Summers, Ed; Waites, William; Young, Jeff; Zeng, Marcia 2011: "Library Linked Data Incubator Group Final Report. W3C Incubator Group Report 25 October 2011." <http://www.w3.org/2005/Incubator/llid/XGR-llid/>

Frosterus, Matias; Tuominen, Jouni; Pessala, Sini; Seppälä, Katri; and Hyvönen, Eero 2013: "Linked Open Ontology Cloud KOKO--Managing a System of Cross-domain Lightweight Ontologies." *The Semantic Web: ESWC 2013 Satellite Events*, pp. 296-297, Springer-Verlag, Berlin Heidelberg, Montpellier, France, May 26-30, 2013.

²⁴ Au sujet des bénéfices du web de données des bibliothèques dans le web de données, voir Baket et al. 2011.

ISO 5964-1985 = International Standards Organization (ISO). Guidelines for the Establishment and Development of Multilingual Thesauri. 1st ed. Geneva: ISO, 1985 (ISO 5964-1985-02-15).

ISO 25964-1:2011 = International Standards Organization (ISO). Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 1: Thesauri for information retrieval. 1st ed. Geneva: ISO, 2011 (ISO 25964-1- 2011-08-05).

Katan, David 2004 (1999): *Translating Cultures: An introduction for Translators, Interpreters and Mediators*. 2nd. ed. Manchester, UK: St. Jemore Publishing.

Masolo, Claudio; Borgo, Stefano; Gangemi, Aldo; Guarino, Nicola; Oltramari, Alessandro 2003: “WonderWeb Deliverable D18. Ontology Library (final).”

Nykyri, Susanna 2010: *Equivalence and translation strategies in multilingual thesaurus construction*. Åbo Akademi University Press. <http://urn.fi/URN:ISBN:978-951-765-521-7>. 413 pp.

Suominen, Osma; Johansson, Alex; Ylikotila, Henri; Tuominen, Jouni; Hyvönen, Eero 2012: “Vocabulary Services Based on SPARQL Endpoints: ONKI Light on SPARQL.” Poster proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2012), Galway, Ireland, October, 2012.

Tuominen, Jouni; Frosterus, Matias; Viljanen, Kim; Hyvönen, Eero 2009: “ONKI SKOS Server for Publishing and Utilizing SKOS Vocabularies and Ontologies as Services.” Proceedings of the 6th European Semantic Web Conference (ESWC 2009), Heraklion, Greece, May 31 - June 4, 2009. Springer-Verlag.

Viljanen, Kim; Tuominen, Jouni; Hyvönen, Eero 2009: “Ontology Libraries for Production Use: The Finnish Ontology Library Service ONKI.” Proceedings of the 6th European Semantic Web Conference (ESWC 2009), Heraklion, Greece, May 31 - June 4, 2009. Springer-Verlag.