

Folksonomies et indexation collaborative

Rôle des réseaux sociaux dans la fabrique de l'information

DocForum

Evelyne BROUDOUX - IUT de Vélizy, UVSQ

Le « web 2.0 » et le tagging

- Un large éventail d'outils dont la pratique centrale est le tagging
- Qu'est-ce que favorise le tagging ?
 - Émergence de groupes, de documents et d'identités d'auteurs
 - Constitution de réseaux sociaux
- Entre expertise et apprentissage : l'apprentissage de l'expertise ?

Qu'est-ce que sont les folksonomies ?

- Historique
 - Terme attribué à Thomas Vander Wal avec l'apparition des services en ligne de type Web2 de partage de ressources par l'indexation (pratique du tagging)
 - Les folksonomies sont des séries de métadonnées créées en collectif par les utilisateurs pour catégoriser et retrouver les ressources en ligne
- Néologisme né de la contraction de Folk + Taxonomy
 - Ne pas confondre les « folksonomies » avec les taxonomies*
 - Taxonomies : classification hiérarchique structurée où chaque terme du vocabulaire est fixé et se range dans un « arbre »
 - Folksonomies : organisation mémorielle de ressources passant par une action de catégorisation - par les utilisateurs - où les relations entre les termes utilisés ne sont pas explicitées

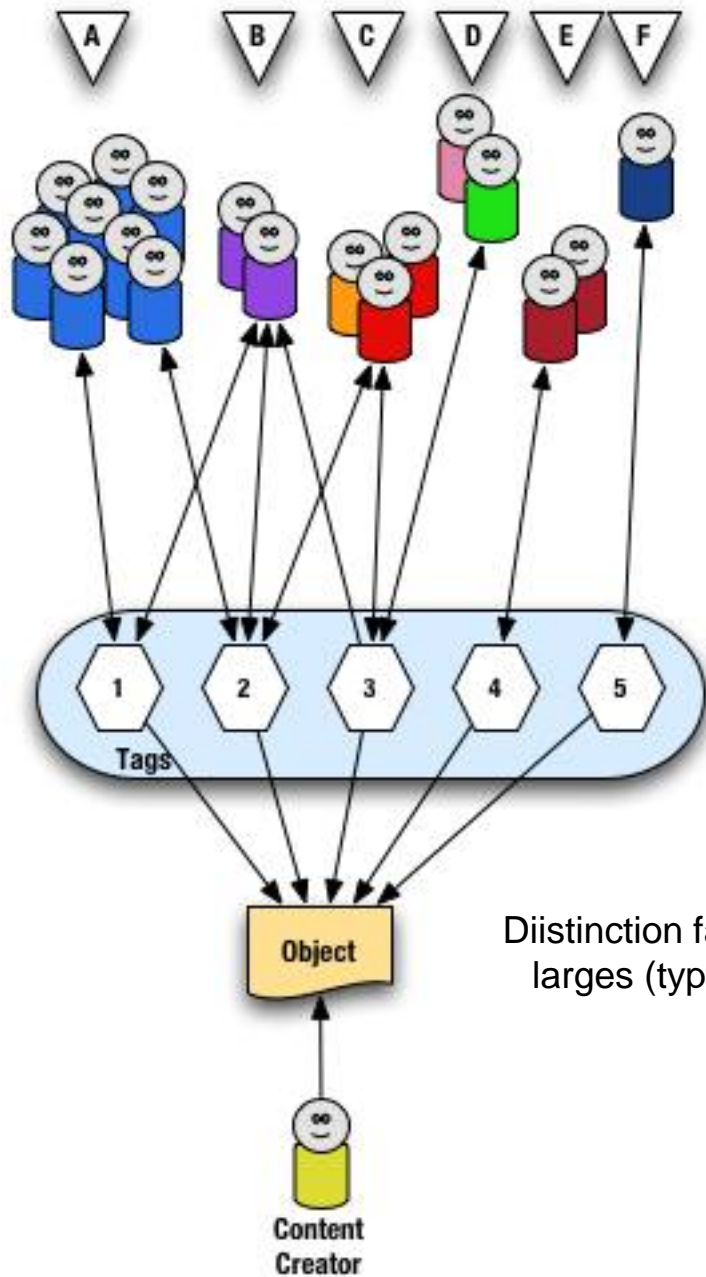
Définition de la folksonomy

- Il faut donc entendre par folksonomy, l'utilisation d'un système flottant de catégorisation, personnel et partagé, permettant de naviguer dans des corpus à large spectre, composés de ressources hétérogènes et instables

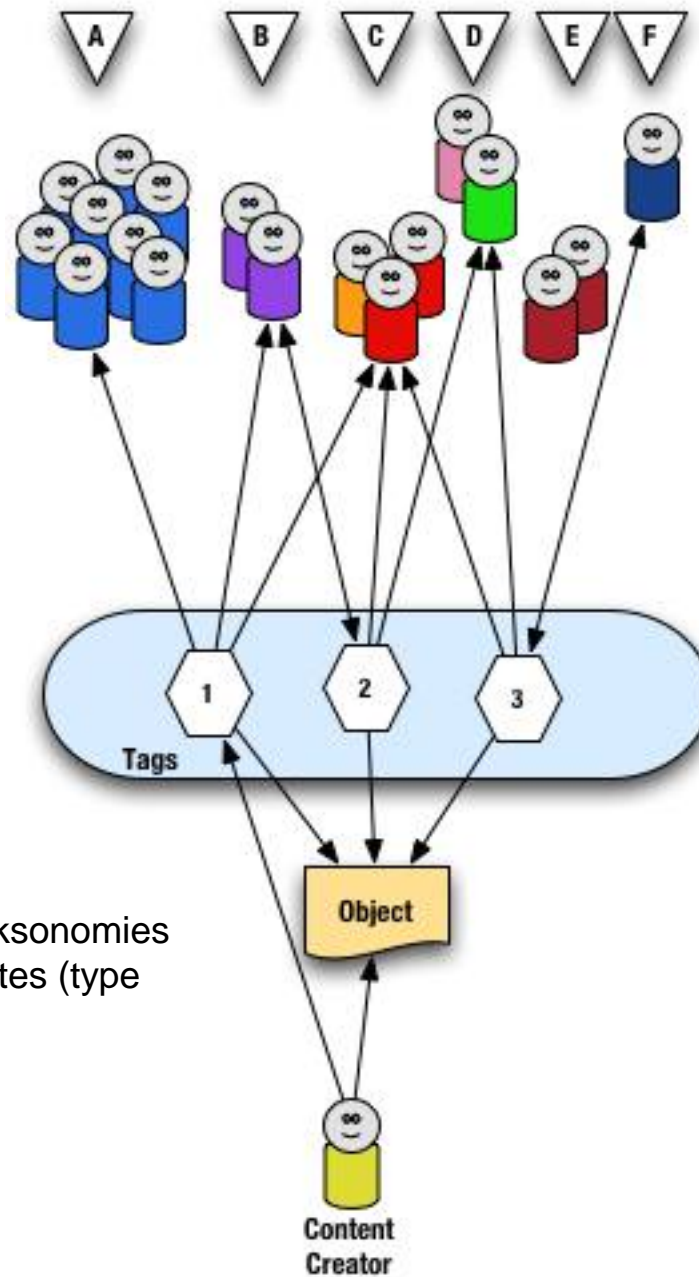
Deux familles d'outils réunis par une pratique

- On peut distinguer deux familles d'outils par la place laissée à l'indexation : celle dont la finalité est l'indexation et celle où l'indexation intervient comme fonctionnalité
- Auto-adaptation : la catégorisation s'effectuant dans un environnement collectif; à tout moment l'utilisateur peut optimiser son système de classement en s'ajustant avec ce que font les autres qui peuvent aussi changer leur manière de faire
- Pratique du taggage
 - Massification de l'indexation par les « amateurs »
 - Ressources
 - Appartenant ou non au taggeur qui les dépose en ligne (création de réservoirs thématiques)
 - Déjà publiées en ligne (taggage comme repérage des ressources existantes)

Broad Folksonomy



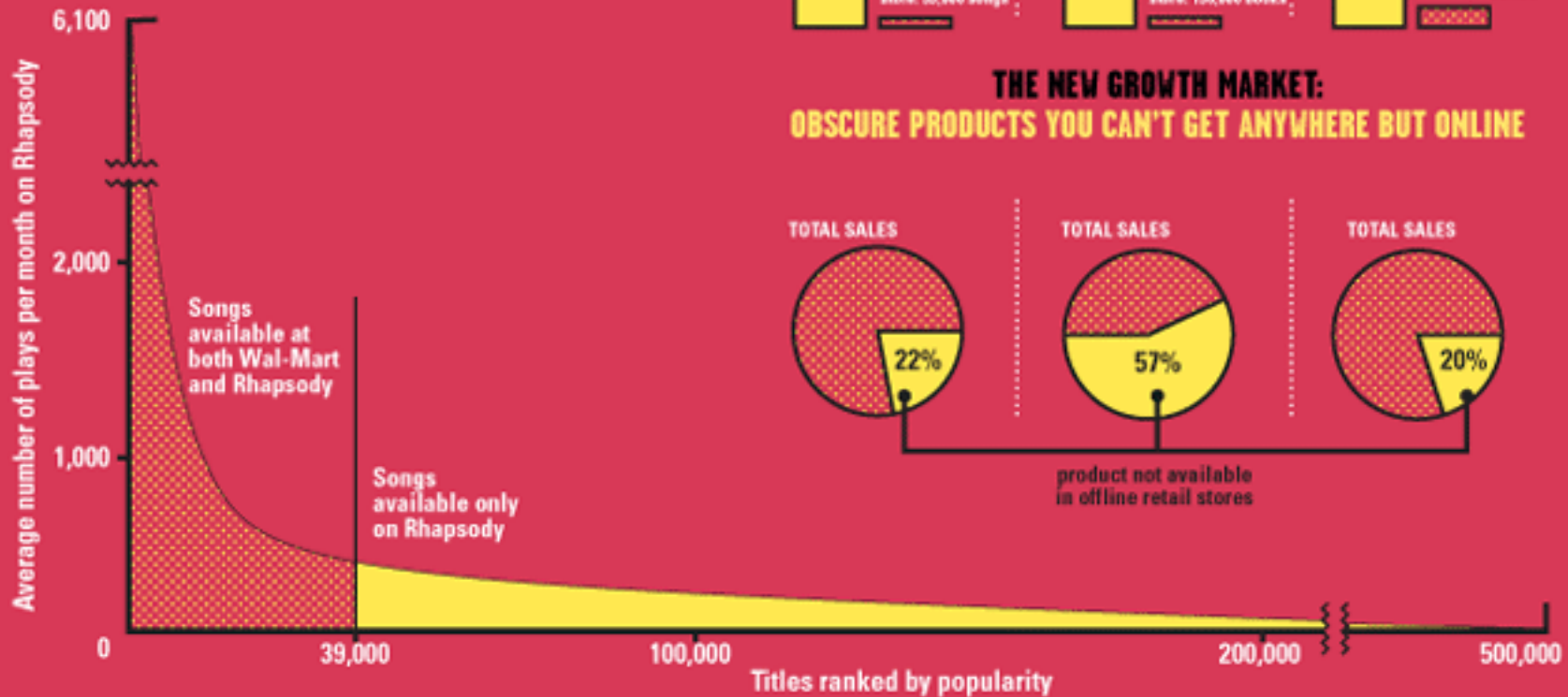
Narrow Folksonomy



Distinction faite par Vander Wal entre folksonomies larges (type Serveur de signets) et étroites (type Serveurs de médias)

ANATOMY OF THE LONG TAIL

Online services carry far more inventory than traditional retailers. Rhapsody, for example, offers 19 times as many songs as Wal-Mart's stock of 39,000 tunes. The appetite for Rhapsody's more obscure tunes (charted below in yellow) makes up the so-called Long Tail. Meanwhile, even as consumers flock to mainstream books, music, and films (right), there is real demand for niche fare found only online.



Sources: Erik Brynjolfsson and Jeffrey Ho, MIT, and Michael Smith, Carnegie Mellon; Barnes & Noble; Netflix; RealNetworks

Folksonomies et reproduction de la « longue traîne »

- Rapport Nbre de ressources/Nbre de taggeurs
 - Une majorité de taggeurs indexe les mêmes ressources => un nombre restreint de tags est utilisé très fréquemment
 - Un volume important de ressources plus « rares » est indexé par un nombre faible de taggeurs => un nombre important de tags est peu partagé
 - Lorsque le contenu est pointu, rares sont les taggeurs
 - Au fur et à mesure, les ressources « rarement » tagguées se diffusent à des cercles s'élargissant
- La folksonomie est-elle vraiment populaire ?
 - Oui, car elle fabrique de la **notoriété**
 - Qu'en est-il du nombre réel de taggeurs ?
 - Le nombre réel de taggeurs construisant des savoirs est réduit et il s'agit d'experts

Fonctionnalités des plate-formes logicielles du « web 2.0 » : créer, repérer, organiser, trouver, partager, communiquer

Production de contenu	Production et évaluation contenu	Gestion de réseau social	Repérage de l'info	Gestion de projets	Publication page web, fils rss	Stockage en ligne	Stockage, publication et gestion de contenus	Indexation Collaborative
Wikipédia								-
		CalendarHub		CalendarHub				CalendarHub
							Flickr YouTube	Flickr YouTube
		Del.icio.us	Del.icio.us					Del.icio.us
	Digg	Digg	Digg					Digg
								Helium
	Helium							
	Memeorandum		Memeorandum					Memeorandum
		MySpace					MySpace	MySpace
			Netvibes	Netvibes	Netvibes			Netvibes
						Openomy		
	Squidoo	Squidoo		Squidoo				Squidoo

Fonctionnalités des serveurs de signets

Fonctionnalités de base	Ajouts de métadonnées sous formes de mots-clés (tags) marquant des URL	Catégorisation : Rangement possible manuel ou semi-automatique des mots-clés par dossier	Commentaires : possibilité de décrire le contenu des URL	Evaluation du contenu	Syndication : balisage en rss du contenu par tag, par utilisateur, par catégorie	Création de dossiers	Distinction Public/Privé
Informations générales de la page d'accueil	Nombre d'utilisateurs actifs, inscrits	Nombre total de mots-clés	Nombre d'utilisateurs par mot-clé	Nombre d'URL tagguées	Aide + Documentation		
Maintenance	Importation/exportation de signets	Modification et suppression du compte					
Opérabilité	Plug-in d'ajouts de signets pour navigateurs	Services additionnels tiers : - nuages tags - visualisation de l'activité	Publication des signets : - blogs - RSS	Technique : - opensource - API - intégration sur sites web			
Recherche	Par mot-clé	Par catégorie	Par utilisateur				
Réseau social	Abonnement aux mots-clés - de tous les utilisateurs (sauf) - par utilisateur	Masquage de la visibilité des abonnements	Création de groupes (suppose un créateur)	Répertoire des utilisateurs	Gestion de l'identité personnelle	Partage de dossiers	Création d'un réseau public/privé
Veille	Abonnement par mots-clés	Génération de fils rss	Abonnement par utilisateur				

QuickTime™ et un
décompresseur TIFF (LZW)
sont requis pour visionner cette image.

QuickTime™ et un
décompresseur TIFF (LZW)
sont requis pour visionner cette image.

Ce que permet le taggage : veille informationnelle

- Création d'une couche signifiante de mots-clés indexant thématiquement des corpus de données, cette couche vient se superposer aux indexations non signifiantes réalisées par les moteurs de recherche
- De la même façon que Wikipédia assemble des contenus produits sans dépôt d'autorité, les folksonomies permettent de créer des corpus de ressources
 - de manière anonyme
 - et qui délimitent des zones d'informations susceptibles de devenir des « savoirs »
- Cette accumulation d'informations se fait en collectif et suivant les outils employés entraîne la création de réseaux publics et (ou) privés

Construction de réseaux

- Morphologie des réseaux
 - Constellation de petits groupes qui peuvent traiter des mêmes thèmes sans se rencontrer = fragmentation
 - Réseaux d'*influence* (qui pointe sur qui ?)
- Un média auto-référentiel
 - Révélé par l'usage décroissant des pseudos
 - Auteur institutionnalisé = utilisation rare
 - Auteur « en devenir » = utilisation courante
 - Instrumentalisé par les bloggeurs cherchant à asseoir leur autorité et augmenter leur notoriété

« [earlier](#) | [later](#) »

[Ruby Java Bridge](#) [save this](#)

by [sylvaincarle](#) to [RoR Java library](#) ... [saved by 17 other people](#) ... 13 mins ago

» [Does your company need a Chief Web Officer?](#) [save this](#)

by [sylvaincarle](#) to [management type:idea](#) ... 14 mins ago

[Atom 1.0 for WordPress](#) [save this](#)

by [sylvaincarle](#) to [wordpress plugin xml atom](#) ... [saved by 37 other people](#) ... 19 mins ago

[Pete Lacey's Weblog :: The S stands for Simple](#) [save this](#)

by [sylvaincarle](#) to [SOA xml webservice type:perspective](#) ... [saved by 398 other people](#) ... 30 mins ago

[rssmarketing.fr » Entretien avec Vincent Maurin Consultant TIC, e-learning, ergonomie, KM](#) [save this](#)

Je l'avais oubliée celle ci...

by [vmaurin](#) to [rss marketing perso interview](#) ... 34 mins ago

[Dabbling With Video Search: A Conversation with Mary Hodder](#) [save this](#)

by [sylvaincarle](#) to [digitalmedia Video search](#) ... 35 mins ago

[Cyberstalking: BlogHers Share Experiences and Advice | BlogHer](#) [save this](#)

by [mathemagenic](#) to [privacy transparency](#) ... [saved by 1 other person](#) ... 37 mins ago

[BBC - Radio Five Live - Pods and Blogs - What's the purpose of tv and radio blogs?](#) [save this](#)

by [mathemagenic](#) to [blogBusiness blogMetrics](#) ... [saved by 5 other people](#) ... 1 hour ago

[Cognitive Edge: Communities of Practice](#) [save this](#)

on [top-down community design](#)

by [mathemagenic](#) to [communities](#) ... 1 hour ago

[Programmer's Notepad](#) [save this](#)

Welcome to [pnotepad.org](#), home of Programmer's Notepad - the free, open source, text editor with special features for coders. (indicates number of characters in the current string as we type)

by [nadc](#) to [notepad editor opensource freeware tools webdev](#) ... [saved by 455 other people](#) ... 1 hour ago

« [earlier](#) | [later](#) »

▼ **spidermm's network (61)**

[20news](#)
♂ [adrien](#)
[alternativa](#)
[aurel](#)
[bhyde](#)
[bushwald](#)
[coolforest](#)
♂ [digitalmonkey](#)
[fanch](#)
[fil.rezo.net](#)
[FrancoisGuite](#)
[fredbird](#)
♂ [GrandCanard](#)
[hrheingold](#)
[Hybernaut](#)
[ittf](#)
[infoworld](#)
[jacques.perconte](#)
♂ [jjejhannet](#)
[joefrontier](#)
[joshua](#)
[judell](#)
[julien](#)
[kase](#)
[kevan](#)
[Linkorama](#)
[makemagazine](#)
[martinlessard](#)
[mathemagenic](#)
[michelmo](#)
[mihaiparparita](#)
♂ [mullingitover](#)
[nando](#)
[nadc](#)
[NetAppRad](#)
[oblondeau](#)
[OF](#)
[overunder](#)
[pement](#)

QuickTime™ et un
décompresseur TIFF (LZW)
sont requis pour visionner cette image.

QuickTime™ et un
décompresseur TIFF (LZW)
sont requis pour visionner cette image.

QuickTime™ et un
décompresseur TIFF (LZW)
sont requis pour visionner cette image.

QuickTime™ et un
décompresseur TIFF (LZW)
sont requis pour visionner cette image.

Folksonomies

« pédagogiques »

- Auto-organisation
 - Indexation des signets : opération de catégorisation
 - Ranger des signets par dossiers suppose des objectifs
- Gains
 - Collecter des informations par domaine d'intérêt
 - Partager ces corpus
 - Créer des réseaux et s'inscrire dans ceux existants
- Risques
 - Informations hétérogènes non évaluées (déperdition d'énergie dû au tri)
 - Instrumentalisation (sectes, marketing « sauvage », relations publiques, etc.)

Processus cognitifs d'acquisition de connaissances

- 3 phases : accroître, structurer, agencer
 - Agrégations de faits et d'informations sur des structures de connaissances existantes (sujets connus, espace-mémoire structuré)
 - Création de structures pour accueillir et interpréter le matériel (nouvelles conceptualisations et créations d'espaces-mémoires)
 - Agencement : contraindre et généraliser les connaissances dans les espaces-mémoires (réarrangement continu)
- Relier des contenus par catégorisation (accroître un domaine de connaissances)
- Structurer à l'aide de dossiers
 - Créer des espaces-mémoires
 - Etablir et remplir des objectifs
- Agencer les dossiers
 - Suppose trier et évaluer les contenus

Organisations concernées

- Création possible de communautés d'apprentissage
 - Par le classement thématique des informations
 - Recherche et structuration
- Communautés de pratiques
- Organisations apprenantes
 - Enseignement
 - Recherche
 - Entreprises
 - (Auto)-Formations

Outils pédagogiques « gratuits » remplaçant certains outils d'assistance à la recherche d'infos (complexes et générateurs de surcharge cognitive)

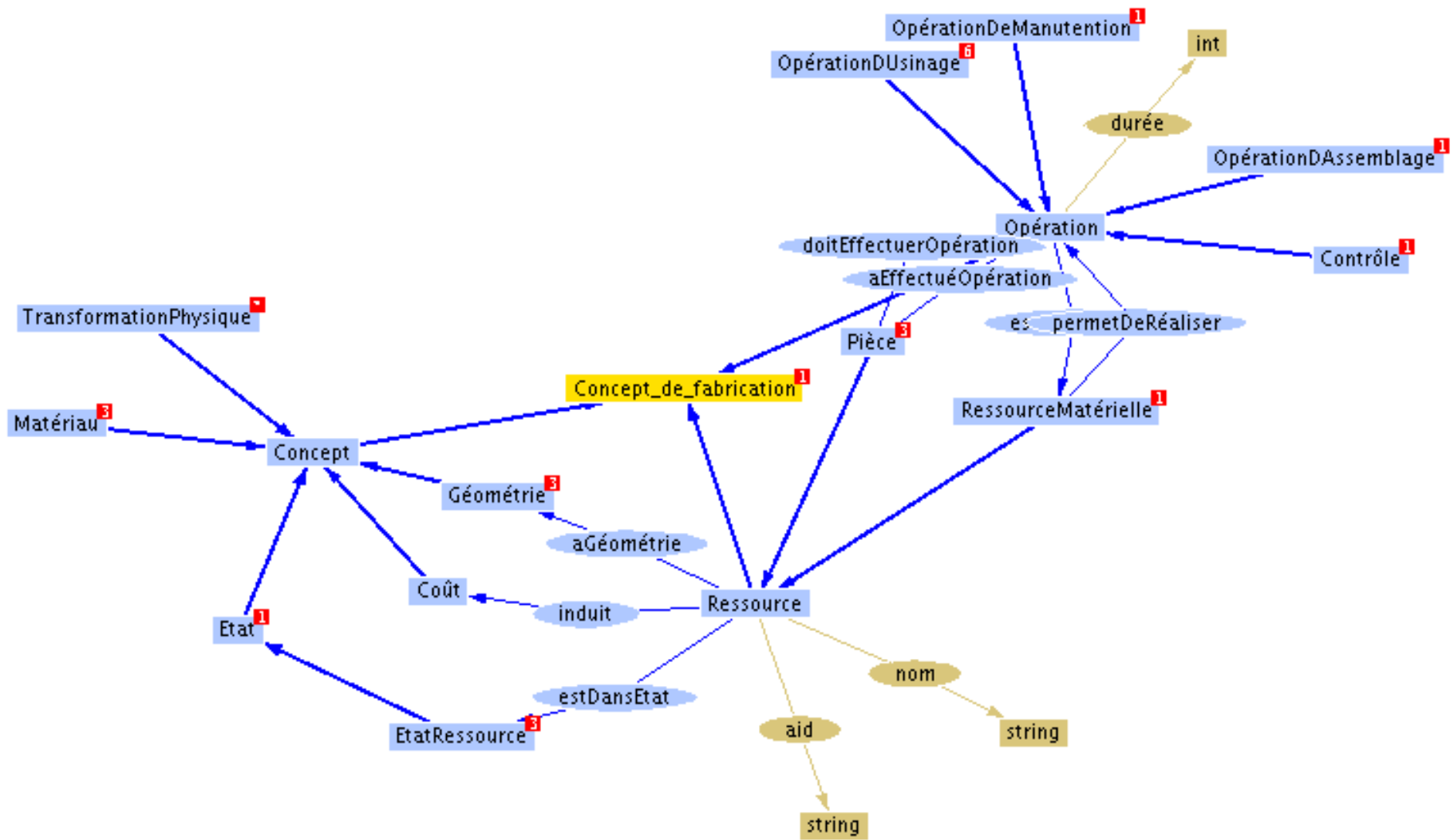
Comment reconnaître et mesurer le degré d'expertise d'un taggateur

- Un domaine d'expertise peut être représenté par
 - Un vocabulaire précis
 - Une évaluation du poids des mots-clés utilisés
- Un domaine d'influence peut être visualisé par
 - L'agrégation de jeux de mots-clés des autres utilisateurs
 - La richesse du réseau de l'indexeur

Taxonomie vs folksonomie vs ontologie ?

- Taxonomie vs folksonomie ?
 - Classement hiérarchique vs rangement de surface
 - Classement exclusif en arbre (un item ne peut appartenir à deux catégories) vs associatif (un item peut être associé à plusieurs tags)
 - Classement d'autorité vs rangement utilisateur
- Ontologie vs folksonomie ?
 - Description de concepts utilisant des catégories formelles et qui suit une structure hiérarchique modulable (graphe)
 - Classement sémantique par héritage des propriétés, un item additionne des significations

Est-ce que les folksonomies sont si « plates » ?



Exemple d'ontologie pour le domaine manufacturier. Source Wikimedia (CC)

Relations hiérarchiques de sujets reconnues dans les folksonomies

- Etude de Kome, 2005
 - Questions
 - Est-ce que l'on peut différencier de grandes catégories à travers le tagging?
 - Est-ce que l'on peut déceler des hiérarchies entre les tags ?
 - Méthodologie
 - 30 URL(s) - 647 posts - 2114 tags individuels
 - Catalogage par 10 étudiants en sciences de l'info et des bibliothèques de la liste d'url avec cochage éventuel de la catégorie « mathématiques » + commentaires
 - Classement alphabétique des tags utilisés par post et recherche d'une relation hiérarchique par groupe de deux tags adjacents
- Réponses positives

Représentation topologique des compétences d'une communauté possible ré-utilisation des « arbres de connaissances »

Arbre : système synchrone auto-organisé et auto-organisateur de représentation de compétences, navigation par proximité

Tronc : premières compétences acquises menant à d'autres apprentissages

Branches : pratiques ou savoirs associés dans la vie réelle des membres d'une collectivité

Feuilles : compétence acquise qui n'est pas nécessaire à l'acquisition d'autres compétences

Feuilles du haut de l'arbre : compétences acquises selon longueur de l'expérience



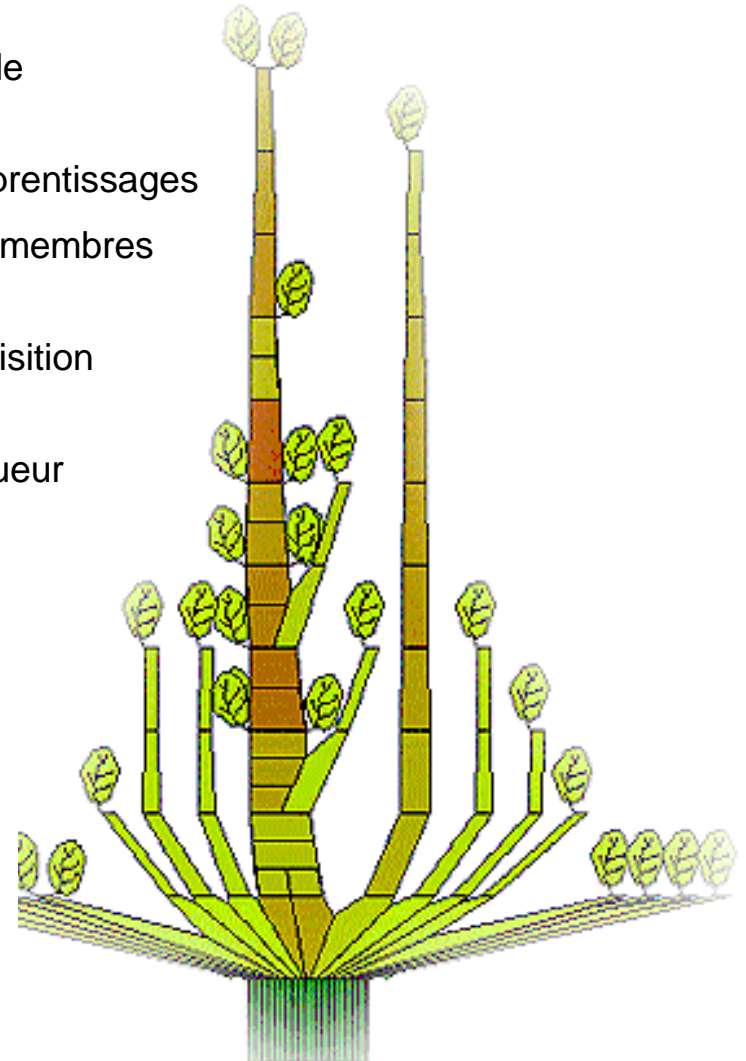
Arbre : représentation d'une communauté apprenante

Tronc : mots-clés partagés par l'ensemble de la communauté

Branches : agrégation de mots-clés révélant des thématiques de recherche (sous-groupes de travail)

Feuilles : corpus délimités par les mots-clés

Feuilles terminant les branches : thématiques expertes nécessitant l'accumulation de savoirs



Outils d'indexation collaborative

- Hier
 - Collectent et partagent des ressources
 - Émettent des avis sur ces ressources
- Aujourd'hui
 - Agrègent des réseaux sociaux
- Demain
 - Évaluation de l'expertise des « taggateurs »?
 - Profilage des usagers avec identifiant unique
 - Poids des réseaux sociaux dans les prises de décision ?

Conclusion

- Pertinence de l'auto-évaluation dans les communautés de pratique
- Croissance et poids des réseaux sociaux
- Outils utiles à l'apprentissage et à l'expertise
- Des usages à imaginer

Bibliographie

- Anderson C. The long tail (<http://thelongtail.com>). Traduction Internet Actu (<http://www.internetactu.net/?p=5911>)
- Gendarmi D., Lanubile F., « Community-driven ontology evolution based on folksonomies », OTM Workshops 2006, LNCS 4277, pp. 181 – 188, 2006.
http://cdg.di.uniba.it/cdg/gendarmi/papers/gendarmi_lanubile_COMINF06_id31.pdf
- Kome H. Sam. « Hierarchical subject relationships in folksonomies ». Master's Paper for the M.S. in I.S degree. November, 2005. 32 pages. Advisor: Jane Greenberg.
<http://hdl.handle.net/1901/238>
- Lévy P., Authier M., « Les arbres de connaissances », Editions La Découverte, 1992. 180 p.
- Mathes A. « Folksonomies - Cooperative classification and communication through shared metadata. Computer mediated communication » - LIS590CMC. Graduate School of Library and Information Science. University of Illinois Urbana-Champaign. December 2004.
<http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
- Rumelhart, D. E., & Norman, D. A. - « Accretion, tuning, and restructuring: Three modes of learning ». In J.R. Anderson (Ed.) Semantic Factors in Cognition (pp. 37-53)

- Liste non exhaustive de serveurs de partage de signets :
 - Blogmarks, Backflip, Blinklist, Blogmarks, Connotea, Del.icio.us, Del.irio.us, FarFetch, Feedmaker, Furl, Jots, Lookmarks, Ma.gniola, Network Menus, NewsVine, Netvouz, RawSugar, Scuttle, Simpy, Spurl, StumbleUpon, Unalog, Wink, Wist, YahooMyWeb