

HDA-Lab... suite

Quelques réflexions sur Wiki/DB-pédia et le tagging sémantique

Bertrand Sajus

Ministère de la Culture et de la Communication
Département des programmes numériques

Sémanticpédia, des données culturelles accessibles à tous
INHA, 19/11/2012





Retour d'expérience

HDA : un corpus qui se prête à un test de tagging sémantique

Anthologie de ressources pour l'histoire des arts

Couvre un champ conceptuel très vaste :

Toutes les périodes

Toutes les civilisations

Tous les domaines artistiques

1er constat : Wikipédia répond aux besoins du tagging "culturel"

Analyse provisoire portant sur les 6 premiers mois de l'année 2012 :

77,2 % des tags ont pu être liés à Wikipédia en français

**83 % des tags utilisés plus d'une fois par les visiteurs du site
Histoiredesarts.culture.fr**

Mois d'octobre 2012 :

**100 % des 100 mots-clés les plus fréquemment utilisés par les
visiteurs du site Histoiredesarts sont liés à Wikipédia**

2^{ème} constat : le tagging par liste de complétion est faisable

Performance technique et expérience utilisateur, un retour positif :

Implémentation de la liste de complétion de Wikipédia

Pas de difficulté technique

Appropriation aisée par l'équipe chargée du tagging

Cette démarche concerne (entre autres) les données documentaires

Méthode applicable a posteriori à des données structurées

La sémantisation du seul thésaurus enrichit la totalité du corpus

8 semaines pour sémantiser 16000 mots-clés



Enrichissement / Augmentation

Le tag augmenté

De multiples formes d'enrichissement :

Désambiguïsation

Gestion de la synonymie

Définitions

Illustrations

Coordonnées géographiques

...

Accès multilingue

Liens entre les différentes versions linguistiques de Wikipédia

Équivalents linguistiques des mots-clés

Adapté aux données non textuelles

Recherche multilingue



Interopérabilité

Interopérabilité sémantique

Wikipédia : pivot du tagging et de l'indexation documentaire

Approche interdisciplinaire

Désenclavement linguistique et culturel des données

Interdisciplinarité : diversité des vues "métier"

Thésaurus de l'Architecture et de l'Urbanisme : Secteur urbain

- **secteur urbain** — 12 thème(s), 60 notice(s) [-]
 - **banlieue** — 1 thème(s), 24 notice(s) [+]
 - **îlot** — 3 notice(s) [+]
 - **quartier** — 1 thème(s), 2 notice(s) [-]
 - **centre-ville** — 2 notice(s) [+]
 - **secteur urbain concerté** — 6 thème(s), 31 notice(s) [-]
 - **grand ensemble** — 8 notice(s) [+]
 - **lotissement** — 4 thème(s), 23 notice(s) [-]
 - **lotissement concerté** — 3 thème(s), 13 notice(s) [-]
 - **cité-jardin** — 7 notice(s) [-]
 - ★★★ Cité jardin, dite cité des Étoiles
 - ★★★ Les jardins historiques et remarquables d'Alsace
 - ★★★ Cité jardin, dite cité des États-Unis ou cité Tony Garnier
 - ★★★ L'oeuvre d'Emile Aillaud en Seine-Saint-Denis
 - ★★★ L'habitat ouvrier dans le pays segréen
 - *Plus de contenus (2)...*
 - **cité ouvrière** — 4 notice(s) [+]
 - **courée** — 2 notice(s) [+]

Thésaurus Garnier : Architecture d'habitation

- **architecture d'habitation** — 18 thème(s), 137 notice(s) [-]
 - **cité-jardin** — 7 notice(s) [-]
 - ★★★ Cité jardin, dite cité des Étoiles
 - ★★★ Les jardins historiques et remarquables d'Alsace
 - ★★★ Cité jardin, dite cité des États-Unis ou cité Tony Garnier
 - ★★★ L'oeuvre d'Emile Aillaud en Seine-Saint-Denis
 - ★★★ L'habitat ouvrier dans le pays segréen
 - *Plus de contenus (2)...*
 - **édifice d'habitation** — 16 thème(s), 130 notice(s) [-]
 - **demeure** — 9 thème(s), 123 notice(s) [-]
 - **chalet** — 5 notice(s) [+]
 - **château** — 45 notice(s) [+]
 - **château-fort** — 2 thème(s), 16 notice(s) [+]
 - **hôtel particulier** — 10 notice(s) [+]
 - **maison** — 24 notice(s) [+]
 - **manoir** — 14 notice(s) [+]
 - **palais** — 9 notice(s) [+]
 - **abri** — 5 thème(s), 7 notice(s) [-]
 - **cabane** — 4 notice(s) [+]
 - **campement** — 3 thème(s), 3 notice(s) [-]
 - **tente** — 1 notice(s) [+]

Désenclavement linguistique et culturel des données

Ex. : Recherche par liste de complétion en français et en japonais
Gothique > Gothique internationale

Français

Rechercher un tag


gothiq

Art gothique	59	Gothique international
Architecture gothique	14	 Le gothique international est une expression inventée au milieu du XIX siècle siècle par l'historien d'art français Louis Courajod pour désigner cette « gothicité universelle » par laquelle les t flamandes, à la fin du XIV siècle
Gothique flamboyant	9	Source : Wikipédia
Gothique internationale	7	
Sculpture gothique	5	
Gothique angevin	5	
Gothique rayonnant	3	
Écriture gothique	1	

Japonais

タグを検索する

ゴシック

ゴシック	59	国際ゴシック
ゴシック建築	14	 国際ゴシック(こくさいい)は、ゴシック美術の末期に、教堂や宮廷を中心に幻想的で細密な絵画が流行した状況に対して、ヨーロッパ共通のスタイルが成立したとして、名づけられたもの。国際ゴシック様式とも。元々は19世紀フランスの美術史家がルネサンスのルーツはフランスにあると主張するために使い始めた言葉のようである(Louis Courajod 1841-1896年)。...
ゴシック・リヴァイヴアル建築	13	
ゴシック建築	9	
国際ゴシック	7	
ゴシック建築	3	
ゴシック建築	1	

ウィキペディア

Accès aux données

Définition contextuelle

Accès à l'article de Wikipédia en japonais



Contribuer

Les enjeux de la contribution à Wikipédia

Le Web des données impose un changement de paradigme

Quid de la "pertinence" ?

Comment donner du sens au déluge des données

Le projet Wikipédia : démarche encyclopédique ouverte

Créer un boucle vertueuse : réutilisation <-> contribution

Rationaliser les efforts de contribution à Wikipédia

Deux entrées :

Quels sujets voulons-nous faire connaître à notre public ?

Quels sont nos sujets les plus recherchés par notre public ?



Conclusion

"L'Encyclopédie a rendu pensable une rupture"

[Question à Roger Chartier]

L'encyclopédie en ligne Wikipédia n'est-elle pas l'aboutissement du projet de Diderot et d'Alembert ?

Dans un sens oui, puisqu'elle repose sur les contributions multiples d'une sorte de société de gens de lettres invisibles. Mais Diderot n'aurait sûrement pas accepté la simple juxtaposition des articles, sans arbre des connaissances ni ordre raisonné, qui caractérise Wikipédia.

**"L'Encyclopédie a rendu pensable une rupture",
Le Monde des livres, 14 janvier 2010.**