

Sujet pour l'ensemble des centres de gestion organisateurs

**ASSISTANT TERRITORIAL DE CONSERVATION DU PATRIMOINE ET DES
BIBLIOTHEQUES PRINCIPAL DE 2^{ème} CLASSE**

**EXAMENS PROFESSIONNELS
DE PROMOTION INTERNE ET D'AVANCEMENT DE GRADE**

SESSION 2014

La rédaction d'une note à l'aide des éléments contenus dans un dossier portant sur la spécialité dans laquelle le candidat se présente.

Durée : 3 heures
Examen d'avancement de grade - Coefficient : 1
Examen de promotion interne - Coefficient : 2

SPÉCIALITÉ : DOCUMENTATION

À LIRE ATTENTIVEMENT AVANT DE TRAITER LE SUJET

- Vous ne devez faire apparaître aucun signe distinctif dans votre copie, ni votre nom ou un nom fictif, ni votre numéro de convocation, ni signature ou paraphe.
- Aucune référence (nom de collectivité, nom de personne, ...) autre que celles figurant le cas échéant sur le sujet ou dans le dossier ne doit apparaître dans votre copie.
- Seul l'usage d'un stylo à encre soit noire, soit bleue est autorisé (bille non effaçable, plume ou feutre). L'utilisation d'une autre couleur, pour écrire ou souligner, sera considérée comme un signe distinctif, de même que l'utilisation d'un surligneur.
- Le non-respect des règles ci-dessus peut entraîner l'annulation de la copie par le jury.
- Les feuilles de brouillon ne seront en aucun cas prises en compte.

Ce dossier contient 23 pages

Il appartient au candidat de vérifier que le document comprend le nombre de pages indiqué

Sujet :

Vous êtes assistant territorial de conservation du patrimoine et des bibliothèques principal de 2^{ème} classe au service de documentation de la commune de Cultureville. À la demande de l'élu en charge de votre secteur, votre directeur vous demande de rédiger à son attention, exclusivement à l'aide des documents joints, une note présentant le web sémantique.

Liste des documents :

- Document 1 :** « Le web sémantique » - http://www.bnf.fr/fr/professionnels/web_semantique donnees/s.web_semantique_intro.html – consulté le 8 décembre 2013 - 1 page
- Document 2 :** « Web sémantique, Open Data et bibliothèques : l'exemple de data.bnf.fr », par Romain WENZ - <http://labobnf.blogspot.fr> - 28 septembre 2012 - 1 page
- Document 3:** « Le territoire de demain sera fondé sur le savoir de ses habitants », par Laura GARCIA VITORIA, Directrice scientifique de la Fondation des Territoires de Demain – <http://lesclesdedemain.lemonde.fr/villes/le-territoire-de-demain-sera-fonde-sur-le-savoir-de-ses-habitants-a-13-144.html> - 11 janvier 2010 - 2 pages
- Document 4:** « Ouverture de la dixième édition de la conférence européenne sur le web sémantique ESWC 2013 : attentes et perspectives », Henri VERDIER, Directeur d'Etalab - <http://www.etalab.gouv.fr/article-ouverture-de-la-dixieme-edition-de-la-conference-europeenne-sur-le-web-semantique-eswc-2013-attent-118103002.html> - 28 mai 2013 - 2 pages
- Document 5 :** « Le Web sémantique : améliorer l'accès aux catalogues des bibliothèques », par Patricia LECHEVALLIER – <http://innovbibliotheque.wordpress.com/2012/05/23/le-web-semantique-ameliorer-lacces-aux-catalogues-des-bibliotheques/www.wordpress> - 23 mai 2012 - 1 page
- Document 6 :** « Web sémantique » - <http://www.enssib.fr/le-dictionnaire/web-semantique> – consulté le 10 décembre 2013 - 1 page
- Document 7 :** « Le web sémantique est un enfermement progressif » - <http://www.agoravox.fr/actualites/technologies/article/le-web-semantique-est-un-113718> – consulté le 16 décembre 2013 - 2 pages
- Document 8 :** « Du Web 2.0 au Web 3.0 » - <http://www.agoravox.fr/actualites/societe/article/du-web-2-0-au-web-3-0-16757> – consulté le 12 décembre 2013 - 2 pages

.../

- Document 9 :** « Enjeux culturels et linguistiques autour des données liées Sémanticpédia et le programme Sémantisation », par Thibault GROUAS, Délégation générale à la langue française et aux langues de France, Journée professionnelle "Les musées à l'heure du numérique : travailler en réseau, réutiliser et contribuer" - http://www.culture.gouv.fr/documentation/joconde/fr/partenaires/AIDE_MUSEES/journee_BDNC_2013/grouas.html - Paris, 7 juin 2013 - 2 pages
- Document 10 :** « Wikimédia France et les musées », par Rémi Mathis, Président de Wikimédia France, Journée professionnelle "Les musées à l'heure du numérique : travailler en réseau, réutiliser et contribuer" - http://www.culture.gouv.fr/documentation/joconde/fr/partenaires/AIDE_MUSEES/journee_BDNC_2013/journee-pres.html - Paris, 7 juin 2013 - 1 page
- Document 11 :** « Le web sémantique, une arme à double tranchant pour le référencement ? », chronique de Laetitia CHESSÉ, Responsable référencement, LINKEO.COM - <http://www.journaldunet.com/solutions/expert/51287/le-web-semantique--une--arme--a--double-tranchant--pour--le--referencement.shtml> - 5 avril 2012 - 2 pages
- Document 12 :** « Avis d'expert : le web sémantique au cœur de la culture et du patrimoine », par Pierre COL, Directeur marketing d'Antidot - <http://www.silicon.fr/avis-dexpert-le-web-semantique-au-coeur-de-la-culture-et-du-patrimoine-87164.html> - 20 juin 2013 - 3 pages

Documents reproduits avec l'autorisation du C.F.C.

Certains documents peuvent comporter des renvois à des notes ou à des documents volontairement non fournis car non indispensables à la compréhension du sujet.

DOCUMENT 1

LE WEB SÉMANTIQUE

Le terme de web sémantique a été proposé par Tim Berners Lee en 2001 (« The Semantic Web », *Scientific American Magazine*, May 17, 2001) pour désigner une évolution du web qui permettrait aux données disponibles (contenus, liens) d'être plus facilement utilisables et interprétables automatiquement, par des agents logiciels. Pour permettre cette évolution, un certain nombre de standards et de technologies ont été développés par le W3C, avec pour objectif de sortir les données des silos fermés que constituent les bases de données en ligne.

Le web sémantique part du principe que les données structurées (par exemple, les métadonnées contenues dans un catalogue de bibliothèque) sont déjà disponibles ; il propose un ensemble de techniques visant à les rendre plus utilisables.

Les technologies du web sémantique permettent de créer de telles données, d'exprimer des vocabulaires et des règles qui les décrivent, et de bâtir des systèmes capables de les manipuler dans de bonnes conditions d'interopérabilité.

Il ne faut pas confondre le web sémantique avec les technologies de traitement automatique des langues (TAL) ou d'intelligence artificielle (IA) qui ont pour objet l'extraction automatique d'information ou de connaissances à partir de données peu ou pas structurées.

http://www.bnf.fr/fr/professionnels/web_semantique_donnees/s.web_semantique_intro.html

DOCUMENT 2

WEB SÉMANTIQUE, OPEN DATA ET BIBLIOTHÈQUES : L'EXEMPLE DE DATA.BNF.FR

Les catalogues de bibliothèque sont des instruments de recherche précis et parfois complexes, en tout cas toujours neutres : ils reprennent des informations factuelles et la censure n'y est pas pratiquée. Le Web les bouscule, du moins dans leur forme traditionnelle. En ligne, les catalogues ne sont pas seulement des outils de gestion et de localisation, mais aussi des outils de référence, pouvant servir à établir une bibliographie, à vérifier une information, ou à se procurer un document numérisé.

Dans le projet data.bnf.fr, lancé en 2011, la Bibliothèque nationale de France organise la rencontre entre les données de ses catalogues et deux secteurs innovants : le Web sémantique et l'*« Open Data »*.

Deux secteurs qui dès les années 2000 ont été perçus comme un seul ensemble : pour désigner la publication de données dans le « nuage » du Web sémantique, on parle de « données ouvertes reliées » (*Linked Open Data*).

Les catalogues de bibliothèques ont justement besoin d'ouverture et de liens. Ils sont composés de données structurées, conçues pour être utilisées dans des programmes informatiques, mais fonctionnent comme des « silos », mis en place avant le développement des standards du Web. Or, les techniques du « Web sémantique » permettent de traiter des informations structurées pour les exploiter dans des programmes automatiques. Pour data.bnf.fr, elles permettent de regrouper des informations provenant des différents catalogues. On pourra par exemple retrouver les différentes œuvres de Voltaire, les éditions du Cid, ou les ouvrages sur l'escrime.

Ce projet utilise en particulier le catalogue général, la base BnF Archives et Manuscrits, et la bibliothèque numérique Gallica. Il s'agit de fournir des fiches sur les auteurs, les œuvres et les thèmes, en regroupant automatiquement les différentes informations. À partir des données structurées des catalogues, on pourra retrouver les différentes représentations d'Hamlet, les événements de la vie littéraire de 1515, ou des photographies de lancer du poids.

À l'heure actuelle, le corpus choisi reflète les auteurs les plus représentés : ceux pour lesquels il existe le plus de documents. Cela permet de générer automatiquement l'ensemble des pages, en s'appuyant sur le travail réalisé par les professionnels du catalogage. En septembre 2012, 3,6 millions de documents sont présents, avec 17 000 auteurs, et il est prévu d'étendre progressivement ce corpus.

Par ailleurs, les informations des catalogues sont produites par des services administratifs, et sont diffusées dans le cadre des missions de diffusion des documents « à distance en utilisant les technologies les plus modernes de transmission des données ». La BnF fait le choix d'expérimenter, en permettant de réutiliser ces données librement à condition de citer la source.

Pour la BnF, la diffusion et la qualité de la structuration sont complémentaires : juridiquement les données sont ouvertes et signalées sur data.gouv.fr, le portail des données publiques de l'État ; techniquement, elles sont disponibles en RDF selon les standards du Web sémantique.

Publié par Romain WENZ

<http://labobnfr.blogspot.fr> - vendredi 28 septembre 2012

DOCUMENT 3

LE TERRITOIRE DE DEMAIN SERA FONDÉ SUR LE SAVOIR DE SES HABITANTS

Par Laura GARCIA VITORIA, Directrice scientifique de la Fondation des Territoires de Demain

Le territoire de demain - notre ville, notre quartier - sera fondé sur le savoir de ses habitants et leurs compétences; notre société sera une société de la connaissance avec tout ce que cela suppose d'échanges, de transferts, de mutualisations.

L'éducation et la formation y auront des formes multiples - tout au long du jour et de la vie. Nos outils de communication et singulièrement ceux de la mobilité en seront les vecteurs privilégiés, tout comme les nouvelles formes de commerce qui se transformeront aux côtés de leurs compléments en ligne.

Nous irons ainsi vers un « monde réel augmenté », où des compléments virtuels de savoirs accompagneront et illustreront les services divers que nous connaissons sur notre territoire. Maints espaces publics seront ainsi des espaces de l'innovation rassemblant usagers et citoyens aux côtés des acteurs économiques et des chercheurs, se faisant tous « passeurs de savoirs » au travers de l'expression de leurs besoins respectifs.

Le développement durable du territoire en sera facilité. Pour prendre le cas du secteur de l'énergie, on sait bien que la consommation de demain, ce n'est pas seulement la production d'énergies renouvelables, c'est d'abord tout un ensemble de réseaux à construire et à interconnecter, c'est une meilleure compréhension surtout, une meilleure intelligence du territoire, qu'il conviendra de mobiliser.

Il ne s'agit pas de vagues projets à un lointain horizon. Des « laboratoires vivants » ont commencé à se développer un peu partout en Europe et dans le monde. Donnons en quelques exemples.

Le Living Lab des Territoires de Demain contribue ainsi notamment à une revitalisation forte d'espaces ruraux un temps délaissés comme en Ardèche - où une communauté de communes entend déployer de nouveaux outils d'attractivité pour les entreprises innovantes - ou encore dans les Hautes Pyrénées - où une autre collectivité territoriale va mettre d'anciens espaces de savoirs à la disposition des savoirs de demain dans une abbaye ouverte à la création numérique et à la réflexion sur de nouvelles formes d'e-tourisme . Il se veut être une préfiguration de tels espaces, dans une Europe ayant retrouvé le sens des engagements de la stratégie de Lisbonne.

DES ESPACES DANS TOUTE L'EUROPE

En Espagne, avec un « technocampus » comme celui en construction à Mataro ou les espaces sociaux d'innovation qui se développent dans l'ensemble de la Catalogne, les quartiers de la connaissance - comme celui de Poblenou à Barcelone - ou encore les maisons du savoir en Estrémadure, les exemples sont multiples de lieux où naissent des formes inattendues d'interaction entre les compétences, où interagissent les volontés de connaître et de comprendre, où les enfants découvrent la robotique, où les savoirs rapprochent les habitants des entrepreneurs les plus dynamiques dans les domaines économiques ou culturels. De même en est-il dans de nombreuses collectivités territoriales finlandaises. Le Web sémantique (une évolution axée sur la recherche de nouveaux fonctionnements pour faciliter le repérage des ressources et la recherche d'informations) est ainsi emblématique de la spatialisation des réseaux de connaissance qui naissent ici et là à

... /

<http://lesclesdedemain.lemonde.fr> – 11 janvier 2010

l'instar des Kompetenznetze allemands, où scientifiques, formateurs et acteurs économiques confrontent leurs analyses et les complètent en temps réel, acquérant par la même de nouveaux statuts, forgeant de nouvelles employabilités pour les décennies à venir et établissant des cartographies de connaissance gérables en temps réel et qui caractériseront chaque moment de nos vies, intégrant par exemple pour chacun des espaces personnels de réflexion prospective.

Ces lieux pour penser demain à l'instar du Centre du Futur au Couvent Saint Sauveur de Venise - des lieux parfois prestigieux, parfois beaucoup plus modestes - seront au cœur de la clustérisation de nos territoires où ce qui comptera plus que les polarités de compétences et de compétitivité seront les liens de coopération qui les unissent. Une économie basée sur la connaissance, ce sera d'abord la démultiplication des liens avec les recherches et les innovations d'autres champs et surtout d'autres territoires. Ainsi se forment un peu partout autant d'écosystèmes ouverts à la créativité - 2009 a été proclamée par l'Union européenne année de la créativité - et surtout - et c'est essentiel - de nouveaux rapports entre grandes, petites et moyennes entreprises.

Au-delà des régions de la connaissance qui se développent au travers de projets européens, de grands espaces de l'innovation apparaissent ici et là en genèse, dans le monde méditerranéen notamment avec le projet EUROMED LAB, ayant vocation à relier les espaces les plus innovants et à en démultiplier la création plus particulièrement sur les rives méridionales, grâce à des institutions internationales, mais aussi à des volontés multiples prochainement rassemblées au Sommet de l'espace numérique méditerranéen à Malaga.

UN ÉQUIVALENT EUROPÉEN DU MIT DE BOSTON

Une telle dynamique qui, mieux que toute autre, définit notre temps est celle des Communautés de Connaissance et d'Innovation qui vont caractériser bientôt le fonctionnement du nouvel Institut Européen d'Innovation et de Technologie (EIT) qui vient de prendre ses quartiers à Budapest. Elle est d'autant mieux partagée à l'échelle mondiale lorsque des responsables du Nord mexicain se mettent à rêver - à l'occasion d'une grande journée de la ville numérique - d'un espace de l'innovation remplaçant la frontière et reliant Chihuahua, Monterrey, Austin et San Diego. Lorsque des habitants du Bénin souhaitent accueillir un «Forum Global Local» soucieux de mettre les compétences de chacun au cœur de la coopération de demain, alors que d'autres au Sénégal entendent créer à 200km de Dakar un grand centre numérique ouvert à tous, ou quand on se met à rêver à Las Palmas d'un monde atlantique de l'innovation reliant trois continents et leur histoire... C'est ce nouvel imaginaire qui nous permettra de construire pour demain une économie et une société durable car basée sur le savoir.

Un dernier mot pour souligner que cette révolution n'est pas faite que de ruptures, mais qu'elle saura préserver les identités. Si les technologies de l'information et de la connaissance y apparaissent comme fournissant de manière privilégiée ces «services de savoirs», ils ne seront pas pour autant un véhicule de mutation purement technologique et de déracinement par rapport à nos pratiques traditionnelles, ils seront tout au contraire de remarquables auxiliaires de la boîte à outils de gestion de nos identités. Souvent en effet dans le passé, nous ne l'oubliions que trop, des territoires se sont transformés en lieux de création et de diffusion de connaissances nouvelles grâce aux hommes qui y habitaient, y travaillaient ou y passaient, forgeant ainsi l'image de régions d'excellence, vrais vecteurs de prospérité. La numérisation des usages et du fonctionnement des institutions s'inscrit dès lors dans autant de continuités que de situations de rupture.

Ouverture de la dixième édition de la conférence européenne sur le web sémantique ESWC 2013 : attentes et perspectives

Tout se passe comme si la révolution numérique voyait naître une troisième phase : la révolution des données, aussi différente de la révolution Internet qu'Internet l'avait été de l'informatique des années 70. C'est la révolution des big data, du web social, de l'open data, ce sont de nouvelles stratégies de création de valeur, de nouveaux codes sociaux, de nouveaux rapports entre les individus, c'est la datavisualisation et les API, le cloud computing, etc.

Avec cette révolution, les entreprises, les citoyens, les administrations ou encore les chercheurs se mettent à travailler au niveau élémentaire de la donnée, à manipuler des distributions complètes, à percevoir des signaux faibles, pour, une fois encore, changer en profondeur l'économie et la société. C'est un tout nouveau rapport au savoir, au pouvoir et à l'action qui se dessine.

C'est dans cette révolution que le web des données prend tout son sens. Le web des données et le web sémantique sont les deux facettes de cette nouvelle infrastructure informationnelle, à la recherche d'un web où **les données de toutes formes produites par tous les acteurs se trouvent interconnectées, identifiées sans ambiguïté et reliées les unes aux autres**. Un web où **la création de nouveaux services, de nouveaux usages, de nouveaux croisements des données est grandement simplifiée** et où la mise en réseau des moyens de production ouvre la voie à **de nouveaux modèles d'innovation**.

La France est engagée dans ce mouvement en tant qu'**acteur à part entière de cette révolution qui voit le numérique connecter le monde**.

Son offre de formations scientifiques de haut niveau, les **compétences des chercheurs français et les travaux scientifiques des laboratoires et instituts de recherche français**, comme le Laboratoire d'Informatique, de Robotique et de Microélectronique de l'Université de Montpellier II (LIRMM), l'Institut National de Recherche en Informatique et en Automatique (INRIA), le Centre national de la recherche scientifique (CNRS) ou le Laboratoire de Recherche en Informatique d'Orsay (LRI), sont largement reconnus à l'international.

La communauté du web sémantique français rassemble **des entreprises performantes et innovantes** – comme Antidot, Atos, Exalead, Logilab, Mondeca, Pearltrees, Semsoft ou Temis – ainsi que des associations comme la FING et le GFII qui diffusent activement les nouveaux usages que permet le web des données.

Les démarches d'**ouverture des données publiques** initiées par l'Etat ou les collectivités territoriales – comme Montpellier avec l'initiative Montpellier Territoire Numérique – contribuent également à l'**émergence du web sémantique**. La diffusion la plus large de données publiques, les vagues technologiques de l'*Open Data* ou du *Big Data*, les nouveaux métiers des *data-scientists* ou des *data-journalistes* contribuent à former une véritable **filière industrielle de la donnée et de ses usages**.

L'Etat lui-même peut aussi bénéficier des avancées issues de la recherche sur le web sémantique. Pour **apporter du sens et rendre interopérables** les données hétérogènes de ses systèmes d'information, l'Etat a notamment élevé sémantiquement des données brutes structurées avec la plateforme open-source du consortium Datalift.

.../

Henri VERDIER, Directeur d'Etalab, <http://www.etalab.gouv.fr>, 28 mai 2013

L'application des technologies sémantiques à l'indexation et à la découverte facilite aussi la **réutilisation des données de l'Etat**. La plateforme française d'Open Data data.gouv.fr est ainsi conçue à partir d'un catalogue sémantique et interopérable de métadonnées descriptives. La Bibliothèque nationale de France (BNF) diffuse par ailleurs sur cette plateforme les fiches de référence sur les auteurs et les œuvres de son catalogue en formats sémantiques. La plateforme Montpellier Données Publiques Ouvertes exploite elle aussi les dernières avancées en la matière.

Ces technologies appuient également la constitution d'**annuaires et de référentiels partagés** au sein des services de l'Etat. L'INSEE publie ainsi des versions sémantiques du Code Officiel Géographique qui recense les subdivisions administratives du territoire ou les données du recensement de la population. Etalab a aussi réalisé, avec le concours de la Direction de l'Information Légale et Administrative (DILA), un référentiel sémantique des administrations de l'Etat.

Le 19 Novembre 2012, le ministère de la Culture et de la Communication a signé une convention avec Wikimédia et l'Inria créant la plate-forme collaborative « **Sémanticpédia** », visant à créer des programmes de recherche et de développement en matière culturelle à partir des données extraites de Wikipédia francophone. La délégation générale à la langue française et aux langues de France soutient également le projet de **version sémantique du Wiktionnaire francophone**. C'est ainsi plus de deux millions de termes qui viendront s'ajouter au réseau sémantique des articles de Wikipédia.

L'ensemble de ces travaux s'inscrivent dans le cadre des études et des projets de la **Commission Européenne** – comme Interoperability Solutions for European Public Administrations, Linked Open Data 2 ou la plateforme ENGAGE – et des **organismes internationaux** tels que le World Wide Web Consortium (W3C).

Etalab, mission du Premier ministre en charge de l'ouverture des données publiques mais aussi de la promotion de l'innovation privée et publique sur ces données, compte s'appuyer sur l'**important travail réalisé par la communauté scientifique française du web sémantique**, et en attend beaucoup pour le développement d'une véritable culture de la donnée en France.

C'est pourquoi je me réjouis de vous assurer du soutien d'Etalab à cette 10^{ème} édition de la Conférence Européenne du Web Sémantique, et souhaite de bons travaux aux intervenants et aux participants qui seront réunis à Montpellier du 27 au 30 Mai 2013.

LE WEB SÉMANTIQUE : AMÉLIORER L'ACCÈS AUX CATALOGUES DES BIBLIOTHÈQUES

À quand le catalogue en ligne de notre bibliothèque municipale qui répondra à la collégienne qui veut "s'occuper de bébés" qu'il existe des ouvrages sur les études de puériculture ? Comment renouveler l'image des catalogues et mieux les inscrire dans les pratiques actuelles des internautes ?

L'accès à distance aux collections des bibliothèques par la consultation des OPAC sera facilité par l'émergence des technologies du Web sémantique. Permettant à la fois une meilleure visibilité des collections et une recherche plus intuitive pour l'utilisateur. Et peut-être à terme d'inscrire ces accès dans l'écosystème du web ?

Grâce à cette innovation technologique : l'ensemble des normes développées par la société W3C structurant les données sur le Web parmi elles le fameux modèle **RDF** (Resource Description Framework ou Cadre de Description des Ressources), sorte de grammaire de base ou langage qui permet aux machines de communiquer et de faire le lien entre les données.

Le décloisonnement de l'ensemble des catalogues sera opéré grâce à des données exploitables et accessibles. La Bibliothèque Nationale de France et le Centre Pompidou adoptent le web sémantique dans la conception de leurs sites Internet pour optimiser et unifier l'accès à leurs différents catalogues. En effet, "*il s'agit de concaténer et de lier la totalité des données provenant de sites divers ou de différents catalogues pour les rendre plus visibles à la lecture*".

Par ailleurs, les **moteurs de recherche sémantique** améliorent la pertinence des résultats par l'**interprétation et l'extension de la requête** avec un résultat plus pertinent, plus exhaustif en évitant le silence par des résultats connexes ou bruités par exemple. Les usagers adopteront en cela une recherche similaire à celle des moteurs de recherche généralistes tels que Google.

Avant d'entreprendre les moyens techniques proposés, il faut faire montre d'une volonté de **transparence sur ses propres données**, comme peut l'expliquer Florence Tisserand dans son mémoire de Master 1 en Information-Communication intitulé "Panorama de la bibliothèque 3.0" afin de correspondre à cette logique de partage avec l'extérieur. Et ainsi attirer le public en offrant un libre accès aux catalogues.

DOCUMENT 6

WEB SÉMANTIQUE

La majorité du Web est destiné à être lu. Il n'est pas fait pour être manipulé de façon intelligente par des programmes informatiques, en général incapables de caractériser les informations qu'ils parcourent.

Le Web sémantique vise à faciliter l'exploitation des données structurées, pour donner du sens au contenu des pages Web, en permettant leur interprétation par des machines. Il ne s'agit pas d'un web à part, mais plutôt d'une extension ou d'une amélioration du Web courant où chaque donnée acquiert un sens défini, afin de créer un réseau d'informations structurées, disponibles en ligne et facilement réutilisables. L'ensemble repose sur des normes, des standards ouverts comme RDF et SPARQL, et des recommandations évitant ainsi redondances, conversions lourdes et permettant la traçabilité des données sources.

Les enjeux en bibliothèques sont multiples. Outre la nécessité d'apporter du sens au web et donc de positionner l'institution dans la société de l'information, la sémantisation du web permet une meilleure interopérabilité et échange des données, leur enrichissement, une exploitation plus fine et une plus grande visibilité répondant, de fait, aux exigences de signalement des collections et de présence en ligne. En tant qu'institutions, elles peuvent apporter une valeur ajoutée certaine dans la validité des données apportées (source, date), la pérennité des accès et citations (ARK, autres identifiants) et leur utilisation par des tiers (tradition d'une offre de services, désintéressement financier).

De fait, parmi les principaux projets, on peut citer le catalogue collectif des bibliothèques suédoises, *Libris*, en 2008, la publication par la Bibliothèque du Congrès de ses autorités sujet (LCSH) en SKOS (Simple Knowledge Organisation Systems ou Système simple d'organisation des connaissances) en 2009, ou encore le projet VIAF (Virtual International Authority File) qui agrège les données d'autorités de plusieurs bibliothèques nationales et dont l'un des objectifs est la publication de ces données dans le Web de données.

En France, les premières expérimentations concrètes dans le domaine des bibliothèques furent l'inclusion de données en RDFa dans le catalogue *Calames* (Catalogue en ligne des archives et manuscrits de l'enseignement supérieur) et la publication expérimentale de RAMEAU en SKOS dans le cadre du projet européen TEL Plus. Depuis 2001, la Bibliothèque nationale de France regroupe toutes les informations issues de ses différents catalogues, ainsi que de sa bibliothèque numérique Gallica, à travers le site data.bnf.fr. Ce n'est donc plus le catalogue qui importe, mais à travers cette ouverture sur le web de données, l'accent est mis sur les ressources (contenus, liens, services forment une unité documentaire).

<http://www.enssib.fr/le-dictionnaire/web-semantique>

LE WEB SÉMANTIQUE EST UN ENFERMEMENT PROGRESSIF

Dans la tarte à la crème que devient la définition d'un web 3.0 sans cesse en devenir, on évoque toujours curieusement l'avènement du web sémantique, ce système dans lequel le réseau comprendrait instinctivement nos demandes.

Point pourtant besoin de carburer à la surconceptualisation et de futuriser ce qui existe déjà dans le temps présent.

Le web sémantique est actuellement l'ensemble de technologies développé par le W3C (l'un des principaux organismes de normalisation du Web que vous pourrez rencontrer au www2012) visant à faciliter l'exploitation des données, en permettant leur interprétation par des programmes et des machines.

Il se décline d'ailleurs dans un usage de plus en plus couru : le web de données. Celui-ci mélange les technologies du Web sémantique avec les principes de base (Identifiants etc.) avec pour objectif la construction d'un réseau d'informations structurées et facilement réutilisables dans de nombreux contextes.

C'est d'ailleurs ce modèle qu'utilisent entre autres deux sociétés aussi classiques aujourd'hui que Facebook et Google.

Cela a été dit de nombreuses fois, Facebook, comme l'est également par exemple Pinterest, est une gigantesque machine à observer et gérer vos comportements.

Les importantes modifications survenues en septembre dernier sont d'ailleurs toutes allées dans le sens d'une collecte de plus en plus grande avec par exemple la possibilité de partager ses goûts musicaux en temps réel sur le réseau via une interconnexion avec Deezer et Spotify. Mais aussi la **création d'outils qui permettent, pour la machine, d'affiner encore sa perception de vos interactions avec les autres** : possibilités de s'abonner aux mises à jour de quelqu'un sans entrer en lien "d'amitié", liste d'"amis proches", etc.

Sur votre fil d'actualité apparaissent, de façon de plus en plus affinée, les informations des gens avec qui vous interagissez le plus par vos commentaires, vos likes, vos listes, etc. et de moins en moins les autres. Du coup, d'outil de communication de réseau plus ou moins vaste, vous passez à la discussion en petit groupe.

Plus vous interagissez avec les uns, moins vous découvrez les autres, plus vous avez d'efforts à faire pour découvrir ce qu'ils ont à dire et à les découvrir en ligne.

Sur Google aussi, la chose existe depuis bien longtemps. C'est même la base du moteur puisque ce sont les liens jugés les plus pertinents par les internautes qui sont privilégiés. Le concept s'est bien sûr renforcé avec la géolocalisation. Tapez le terme "restaurant" depuis un ordinateur situé à Lyon et sur un autre à Tours, les résultats ne seront pas les mêmes. Google plus a encore renforcé cet aspect, en permettant à ses usagers de recommander à leurs contacts des résultats de recherche. Les liens ainsi recommandés amènent à ce que l'internaute perçoive en priorité les résultats de ses cercles "d'amis"...

.../

<http://www.enssib.fr/le-dictionnaire/web-semantique>

De plus en plus, cette dimension sémantique prend le pas. **C'est parfois fort utile pour trouver exactement ce que l'on cherche mais risque à terme de nous enfermer dans un simple cocon**, freinant là aussi notre possibilité de découvrir, mettant le mot sérendipité dans le rayon poussiéreux d'un passé révolu.

Ce risque de fragmentation sociale online existe : reprenons le cas de Google et des moteurs de recherche en général. Leur fonctionnement optimal (et leur business modèle de publicités personnalisées) est basé sur le fait d'arriver à trouver au plus facilement ce que veut celui qui les utilise et éliminer ce qui est non pertinent.

Dans cet esprit, se pose la question de l'apprentissage de la machine. Prenons le cas d'un internaute qui tapera "Elysee" sur google en cherchant non la résidence du Président de la république mais le théâtre du même nom. Au départ, la machine propose d'abord comme résultat le plus courant et le mieux référencé des résultats donc, celui du Palais présidentiel. **Mais au bout de plusieurs reprises où l'internaute aura en lieu et place cliqué sur le lien du site du théâtre, la machine le lui proposera alors spontanément en premier résultat.** Voilà un formidable défi pour les références -qui est d'ailleurs aussi un moyen pour Google de rendre ses publicités encore plus indispensables en marginalisant un peu le référencement naturel-.

C'est aussi, sous des abords pratiques (plus besoin de fouiller dans les résultats pour trouver le théâtre de l'Elysée ou qu'on me propose des restaurants tourangeaux en lieu et place d'établissements situés à Lyon) le risque d'enfermement dans une bulle, fin partielle de cette promesse d'ouverture généralisée que promet pourtant le web.

DOCUMENT 8

DU WEB 2.0 AU WEB 3.0

Comme en témoignent de nombreux articles et sondages, la compréhension de ce que recouvre le Web 2.0 est plutôt mitigée. Un grand nombre de personnes interrogées ne savent en effet pas ce que c'est, tandis que d'autres n'y voient qu'un phénomène de mode.

Malgré ce flou, on commence déjà à parler du Web 3.0, la prochaine génération d'Internet, qui apporterait de nouvelles possibilités et davantage de services. La conférence Le web 3 organisée par Loïc Le Meur témoigne de ce phénomène. Analyse.

Souhaitant apporter un éclairage sur cette récente tendance, j'ai tenté d'expliquer de façon synthétique les apports successifs des différentes générations du Web par la cartographie ci-jointe. Cette analyse n'est vraisemblablement pas exhaustive, et très certainement incomplète, mais elle a pour vocation de fixer un certain nombre de repères essentiels pour comprendre les étapes majeures de l'évolution Internet. Elle apporte également des éléments sur ce que sera le Web dans quelques années.

J'ai retenu cinq axes : la technologie, le contenu, le commerce, le pouvoir et l'influence, la dimension sociale.

- Technologie : *la véritable source de progrès, sans laquelle rien n'est possible*

La figure présente quelques innovations majeures, allant des premières pages HTML, au Web sémantique, en passant par le XML, les flux RSS, et la technologie AJAX.

Non, il ne s'agit pas d'une grande lessive (!), mais bien d'évolutions majeures, apportant à chaque étape des possibilités nouvelles. Le Web 3 devrait donc être synonyme de Web sémantique, permettant par exemple d'obtenir une réponse cohérente et précise, à une question ciblée. Cela devrait être possible en associant au contenu une couche de description, c'est-à-dire en utilisant des informations sur les informations (les métadonnées).

- Contenu : *grâce aux progrès de la technologie, le contenu s'améliore*

A la fin des années 1990, le contenu était fourni par des sites Internet présentant des pages HTML. La seconde génération Internet (Web 2.0) a enrichi ce modèle par des documents multimédias connectés entre eux. Moins « texte » et davantage « vidéo » et « son », le nouveau Web franchit ainsi une nouvelle étape. En même temps, les échanges et le partage se développent. La révolution attendue du Web 3.0 consiste en une plus grande intégration des sources d'information, avec des réponses à nos requêtes plus intelligentes, plus pertinentes, et plus proches de l'humain. Bref, l'ère du Web 3 sera marquée par la déduction cognitive.

- Commerce : *vers une intégration multi-acteurs et multimédia*

Le Web 1 a créé véritablement une révolution dans le commerce, d'une part par l'arrivée de marchands Web tels qu'Amazon, d'autre part grâce au développement sur Internet du commerce traditionnel (Fnac, SNCF, etc.).

Depuis quelques années, on s'oriente peu à peu vers une plus grande intégration des différents acteurs, qui proposent des services connectés et multimédia.

.../

<http://www.agoravox.fr>

On peut imaginer que demain, l'évolution technologique permettra au commerce en ligne une parfaite intégration entre différents acteurs plus spécialisés et participant à la chaîne de vente (vente, prélèvement du produit de son lieu de stockage, paiement, expédition, facturation). L'intégration signifiera une parfaite transparence pour l'acheteur, qui pourra passer commande et suivre avec facilité sa trace depuis un ordinateur, un mobile, ou tout autre périphérique portable.

- Pouvoir et influence : le pouvoir et les cercles d'influence sont modifiés

Les débuts d'Internet étaient caractérisés par des sites d'information et/ou marchands, auprès desquels les internautes allaient chercher de l'information. Le pouvoir d'influence était la plupart du temps entre les mains de sites professionnels. Avec l'arrivée des outils de partage de l'information (blogs, wikis), le pouvoir est désormais rendu aux utilisateurs. A côté des journalistes professionnels, des individus blogueurs délivrent désormais de l'information, souvent appréciée pour son authenticité et son style direct. Cette tendance devrait s'accentuer avec le Web 3.

- Dimension sociale : vers une participation élargie des individus

Au-delà du pouvoir et de l'influence, il est intéressant de noter que les individus participent de plus en plus à la vie publique, notamment *via* les blogs. Nous sommes passés d'une étape de publication d'informations (grâce aux pages perso) à une phase d'échange. Le Web 3 renforcera à coup sûr le phénomène collaboratif, où le contenu comptera davantage que le statut.

DOCUMENT 9

ENJEUX CULTURELS ET LINGUISTIQUES AUTOUR DES DONNÉES LIÉES : SÉMANTICPÉDIA ET LE PROGRAMME SÉMANTISATION

Les technologies numériques concourent grandement aux objectifs de la politique de la langue du ministère de la Culture : préserver la langue française pour permettre de répondre aux besoins d'expression et de communication des citoyens et des institutions d'une part, favoriser le passage entre le français et les autres langues afin de favoriser le multilinguisme, d'autre part.

La Délégation générale à la langue française et aux langues de France, en raison de son positionnement interministériel et de sa vision d'ensemble des enjeux linguistiques propres au développement des technologies numériques, est en mesure de jouer un rôle de catalyseur. Ce positionnement doit lui permettre d'orienter son action autour de trois priorités : prendre en compte la dimension linguistique des technologies numériques, contribuer à mettre celles-ci au service de la politique de la langue, veiller à la présence de la langue française sur la Toile et aux moyens qui l'encouragent.

Pour mener à bien cette politique en faveur du numérique, le ministère accompagne différents types de projets en faveur du français et de la diversité linguistique :

1. Sur les technologies de la langue

Les technologies de la langue constituent pour les pouvoirs publics une question d'intérêt général : elles contribuent à l'amélioration de la vie quotidienne de nos concitoyens, au développement de notre économie et au renforcement de nos échanges.

Ces technologies permettent de multiples applications telles que :

- la traduction automatique et l'aide à la traduction ;
- l'aide à la rédaction (correcteurs orthographiques et grammaticaux) ;
- la reconnaissance vocale et les commandes vocales ;
- la synthèse vocale ;
- l'indexation automatique de documents ;
- l'automatisation des processus de constitution de méta-données (résumés, mots-clés, catégories ...).

2. Sur l'Internet et les médias collaboratifs

La présence du français sur l'Internet passe notamment par une participation active de nos concitoyens aux différents projets collaboratifs. Il est donc particulièrement important pour le ministère de s'assurer du dynamisme de l'effort collaboratif en langue française.

À ce titre, deux aspects apparaissent particulièrement stratégiques :

- l'enrichissement des encyclopédies ou dictionnaires collaboratifs en ligne ;
- la traduction en français des logiciels, sites internet, réseaux sociaux et outils proposés aux utilisateurs.

3. Sur le Web des données et le web sémantique

C'est un enjeu culturel majeur de développer en français des outils de diffusion culturelle qui confortent le rôle historique de la langue française comme langue internationale de diffusion des savoirs.

Grâce aux nouvelles technologies du web sémantique, l'Internet est en passe de devenir une base de connaissances mondiale. Ce que l'on appelle le "web de données" consiste en effet à interconnecter d'immenses référentiels (terminologies, catalogues d'œuvres...) ouvrant la voie à des usages radicalement nouveaux des données numériques.

.../

Journée professionnelle "Les musées à l'heure du numérique : travailler en réseau, réutiliser et contribuer" - Thibault GROUAS, Délégation Générale à La Langue française Et aux langues de France - <http://www.culture.gouv.fr> - Paris, 7 juin 2013

Le ministère de la Culture, souhaitant que la langue française soit au cœur du web des données, a lancé en novembre 2012 un partenariat stratégique dénommé Sémanticpédia avec INRIA et l'association Wikimédia France, afin de développer l'écosystème naissant autour des données culturelles liées.

Le premier projet développé dans ce cadre a été l'extraction semi-automatisée des données structurées de Wikipédia en français et des langues qui y sont liées. Ce projet, utilisable gratuitement et librement sur internet et dénommé DBPédia en français, s'appuie sur le projet international DBpedia.org. Il a été développé par l'Inria et permet d'ajouter de manière simple des données en provenance de Wikipédia sur des sites ou services proposés par des acteurs publics ou privés.

Le projet HDA Lab, qui reprend les données du site portail Histoire des Arts en les croisant avec celles issues de Wikipédia via DBpedia, est un des premiers projets de ce type mené au ministère de la Culture.

Le programme « Sémantisation », inscrit au schéma directeur des systèmes d'information 2013-2015 du ministère et au programme ministériel de modernisation du ministère (PMMS), vise à tirer les bénéfices de cette nouvelle méthode de travail et de publication sur internet, qui permet l'enrichissement des contenus du ministère avec d'autres contenus tiers, tels ceux présents sur la base DBpedia en français.

Il facilite par ailleurs le développement d'interfaces de navigation multilingues sans nécessiter de traduction, ce qui ouvre la voie à une diffusion plus large des contenus culturels français dans le monde.

Dans le cadre du programme « Sémantisation », plusieurs projets sont envisagés :

- Sémantisation de la base Joconde, gérée par le service des musées de France (SMF) et comptant plus de 300 000 notices illustrées (2013). Ce projet proposera des innovations de l'interface permettant la navigation et l'accès à l'image dans plusieurs langues, et pourra préfigurer l'interface des prochains outils utilisés au ministère pour enrichir et alimenter les bases de données culturelles : MISTRAL et GINCO notamment.
- Sémantisation du Wiktionnaire, en lien avec la communauté de linguistes et chercheurs en traitement automatisé de la langue (2014-2015)
- Sémantisation du Corpus de la parole, base de données opérée par la DGLFLF et contenant plus de 1000 extraits sonores documentés en diverses langues de France (2014-2015).

En termes de réutilisations innovantes, il est par ailleurs proposé de travailler sur le développement d'applications s'appuyant sur DBpedia en français et/ou sur les données sémantisées de Joconde.

C'est dans ce contexte qu'est développée l'application pour terminaux mobiles Muséophile, dont l'objectif principal est de démontrer la faculté de développer rapidement (2 mois) et à moindre coût une application multilingue (8 langues proposées) donnant accès à un corpus de données large et multilingue à partir d'une interface utilisateur simple et ergonomique.

DOCUMENT 10

WIKIMÉDIA FRANCE ET LES MUSÉES

Rémi MATHIS, Président de Wikimédia France

Les moyens de mettre nos collections à disposition du public et de les faire connaître du plus grand nombre ont radicalement évolué en quelques années. Parallèlement, les demandes de ce public ont également connu une mutation franche, dont nous devons tenir compte.

Wikipédia est ainsi devenu en une dizaine d'années le premier site culturel français, avec 20 millions de visiteurs uniques par mois. Il est donc de toute première instance que les institutions y aient leur place, et qu'elles puissent profiter de l'existence d'un tel site pour y atteindre un public qui serait difficilement touché par les moyens traditionnels. Le portail Europeana rappelait ainsi dans sa dernière lettre d'information que quelques images versées sur Commons - la médiathèque dans laquelle Wikipédia va chercher ses images - ont été vues un million de fois, au prix d'un petit travail éditorial internet externe.

Certaines institutions ont commencé à verser le contenu qu'elles produisent. La Ville de Toulouse a signé un partenariat avec Wikimédia France dès 2010 (en même temps que la BnF et le château de Versailles, sur d'autres thématiques), ce qui permit à des documents numérisés par la bibliothèque et les archives d'être présents, mais aussi à des photographies des collections du Muséum, d'être valorisées sur Wikimedia Commons et Wikipédia. Une photo du Muséum récemment élue "Meilleure photo 2012" sur Commons a joui d'une couverture presse internationale. Plus récemment, les musées départementaux de la Haute-Saône ont décidé de verser les images produites pour Joconde : moyen pour eux de gagner en visibilité, de se trouver là où sont déjà les personnes intéressées par les sujets traités... mais aussi de réduire les coûts en externalisant l'hébergement de ces images.

Les images versées avec l'accompagnement de Wikimédia France sont accompagnées des descriptions (métadonnées) produites par les musées eux-mêmes, afin d'assurer une qualité maximale grâce au professionnalisme du personnel de ces institutions. Les données qui n'appartiennent pas déjà au domaine public sont obligatoirement placées sous une licence libre (Creative Commons By-SA), qui permet une réutilisation maximale – ce qui bénéficie en premier lieu aux institutions elles-mêmes, qui peuvent partager ces métadonnées (pour cataloguer son propre fonds d'estampes, par exemple) et réutiliser les images (comparaison sur son site Internet, enseignement, catalogues...).

Cette ouverture massive est déjà largement en œuvre aux Etats-Unis, Allemagne, Pays-Bas, etc. : c'est aussi la place de la culture française sur l'Internet français qui se joue aujourd'hui puisque les grandes périodes de l'histoire et de l'histoire de l'art sont pour l'instant illustrées à partir des collections étrangères, y compris pour les plus grands artistes français

Journée professionnelle "Les musées à l'heure du numérique : travailler en réseau, réutiliser et contribuer" - <http://www.culture.gouv.fr> - Paris, 7 juin 2013

LE WEB SÉMANTIQUE, UNE ARME À DOUBLE TRANCHANT POUR LE RÉFÉRENCEMENT ?

La recherche sémantique sur Google échauffe les esprits des référenciers. En déclarant dans le Wall Street Journal qu'à moyen terme 20% des résultats de recherche devraient être impactés par la recherche sémantique, un haut responsable de Google a affolé les éditeurs web.

Pour rappel, la recherche sémantique, c'est tout simplement la capacité d'un moteur de recherche à comprendre ce dont parle une page, ou ce que recherche un internaute et de fournir directement une réponse pertinente aux requêtes qui lui sont envoyées. Par exemple, dans un futur proche, en tapant « Score Nadal Federer Roland Garros 2008 » sur Google, vous devriez voir s'afficher directement « 6-1, 6-3, 6-0 » sans même avoir besoin de consulter un site... score que Google aura été récupérer lui même sur la fiche Wikipédia correspondante, ou sur le portail de Roland Garros, ou plus vraisemblablement en croisant plusieurs sources d'informations pour s'assurer de l'exactitude de la réponse. Ce n'est pas pour tout de suite, mais quand même...

Google expérimente cette approche depuis quelques années déjà sur des thématiques bien précises. Par exemple, lorsque vous tapez 5*4 dans le moteur de recherche, ce dernier renvoie le résultat de l'opération. Ou lorsque vous tapez « meteo rennes » il affiche les prévisions météo au lieu de vous renvoyer sur meteofrance.fr.

Cette nouvelle n'a pas forcément été bien accueillie dans le milieu du SEO et des éditeurs de site. En effet, si cette innovation va simplifier la vie des internautes, ce n'est pas forcément une avancée pour les éditeurs de site qui fourniront de l'information à Google mais n'en retireront aucun trafic !

Mais n'oublions pas que Google est encore et toujours un robot qui ingurgite, analyse et restitue les informations que l'on veut bien lui fournir.

Et si le Web sémantique était une opportunité pour le référencement ?

Plus globalement, on assiste sur la toile à la naissance progressive du web sémantique. L'objectif du web sémantique est d'organiser le web en donnant du sens à chaque contenu. Plusieurs formats de balisage sémantique existent même si Google, Yahoo et Bing se sont accordés pour recommander les « micro-données ».

En utilisant des micro-données dans leurs pages web, les éditeurs de sites permettent ainsi aux moteurs de recherche -ou à n'importe quel autre service- d'extraire des données sémantiques en rapport avec leur contenu (nom de l'entreprise, téléphone, adresse, nom d'un produit, marque etc.).

L'ensemble des données interprétable, pour le moment, par google est disponible ici : <http://support.google.com>

.../

À l'heure actuelle, Google lit les micro-données présentes sur les sites et les utilise (entre autre) pour afficher des informations supplémentaires dans ses « rich snippets ». Si vous l'ignorez, les rich snippets sont les informations supplémentaires qui apparaissent sous certains résultats de Google (adresse, carte, avis, note, tarifs, etc.).

Prenons un exemple. Sur un portail de réservation en ligne d'hôtels, la fiche d'un établissement peut contenir les balises suivantes : prix, adresse, note donnée par les internautes et avis. Lorsqu'un moteur de recherche, ou plus largement un robot, analysera cette page pour l'indexer, il comprendra immédiatement que "55 euros" correspond au prix de la chambre, que "3 chemin des lanternes, 76000 Rouen" est l'adresse de l'hôtel, et que « 4,5/5 » est la note attribuée à cet hôtel par les internautes.

Regardons maintenant comment Google utilise ces informations dans ses rich snippets.

En tapant « hotel des phares » sur Google, en plus du site « officiel » de l'hôtel, deux sites de réservation en ligne se distinguent parmi les 10 premiers résultats naturels.

En utilisant les micro-données pour qualifier certaines informations comme le prix ou les avis consommateurs, on voit clairement que ces deux sites bénéficient d'une visibilité accrue sur la première page... Peu importe leur classement, ils auront probablement pour cette raison un taux de clic largement supérieur à tous les autres sites présents sur la même page de résultat.

En utilisant ces balises de formatage, on peut maintenant facilement ajouter une note à sa page web et la mettre en avant dans le listing des résultats de recherche. De plus en plus de sites de vente en ligne utilisent aujourd'hui ces « ratings » pour sortir du lot et développer leur visibilité à moindre coût.

De manière générale, le web sémantique est plutôt une opportunité pour les webmasters de qualifier leurs données. Dans les années à venir, ce type de formatage sera certainement un des critères préféré de Google et des moteurs de recherche pour juger de la pertinence d'un contenu.

Comme la balise « keyword » en son temps, il y a fort à parier qu'il y aura moyen de jouer sur ces données pour gagner en pertinence auprès de Google. Reste simplement à espérer que ce dernier n'abusera pas de sa position dominante pour afficher des résultats de recherche de plus en plus riches et garder les internautes sur son site...

AVIS D'EXPERT : LE WEB SÉMANTIQUE AU CŒUR DE LA CULTURE ET DU PATRIMOINE

Pierre COL, directeur marketing d'Antidot, nous parle aujourd'hui de l'utilisation des technologies du web sémantique dans le cadre de la mise en valeur du patrimoine culturel français sur la Toile.

Le web s'est peu à peu imposé comme un vecteur de diffusion incontournable du patrimoine culturel. Les initiatives de numérisation et de mise à disposition via un site web de la richesse du patrimoine ont permis de prolonger sur Internet les actions des institutions de ce secteur (musées, archives, bibliothèques...) et de renforcer la diffusion de la culture au plus grand nombre dans un nouvel espace sans frontières.

Pour autant il est permis de se demander si, dans cet espace où la navigation se fait librement, la logique institutionnelle a encore un sens. En effet, la consultation des informations culturelles sur le web ne part pas d'une action délibérée qui implique la visite d'un site précis, mais d'opportunités en rapport avec des recherches personnelles ou liées à des études ou en relation avec un intérêt ponctuel ou régulier.

Ce mode de « consommation » de la culture et de son patrimoine nécessite de provoquer la sérendipité, c'est-à-dire de favoriser le fait de trouver au hasard une information intéressante. Or de plus en plus l'internaute privilégie certaines grandes portes d'entrée sur le web que sont les moteurs de recherche comme Google, Bing ou Yahoo !, les réseaux sociaux Facebook, Twitter etc., ou encore Wikipedia.

C'est en partant de ces grands sites qu'il va pouvoir rebondir de lien en lien pour atteindre la ressource culturelle qui l'intéresse, ce n'est pas en allant directement sur le site d'une institution culturelle. En d'autres termes, un internaute ne décide pas a priori d'aller visiter le site web du Musée d'Orsay : il veut avant tout découvrir les tableaux de Monet et lire des informations intéressantes sur ce peintre. Et pour cela, généralement, il commence par chercher « Monet » sur Google.

Il est donc plus que jamais nécessaire de mettre en relation les différentes organisations du monde de la culture et de systématiser la création de liens entre les différentes collections pour faciliter la navigation et, donc, la découverte de l'utilisateur.

Pour autant, ces nouvelles formes de médiation ne remettent pas en cause l'existence des institutions qui trouvent leurs racines dans l'histoire de la construction du patrimoine et aussi dans des logiques de fonctionnement différentes : les livres mis à disposition dans les bibliothèques, les archives ou les pièces conservées dans les musées sont de natures très différentes.

.../

Qu'apporte le web sémantique au monde de la culture et du patrimoine ?

La question est donc de savoir comment réconcilier à la fois la logique de liens au cœur du web et la logique des collections et des institutions. Pour ce faire, les conditions d'interopérabilité doivent être réunies, à savoir :

- un protocole de transport commun, HTTP ;
- des standards d'échange d'informations, XML ou JSON ;
- une structure commune avec les technologies du web sémantique mises au point depuis 10 ans au sein du W3C, l'organisme de normalisation du Web.

Les avantages de l'approche du web sémantique sont multiples :

- désigner chaque entité par un identifiant universel et pérenne, qu'on peut réutiliser dans des contextes différents ;
- mieux qualifier les entités décrites dans une notice en les identifiant explicitement, structurer plus clairement certaines informations et mettre en relation les bonnes informations internes ;
- mettre en relation des informations internes avec des informations externes : en effet, le modèle RDF se diffuse dans le monde de la culture (cf. data.bnf.fr ou data.europeana.eu), et de nombreux ensembles de données sont déjà disponibles en RDF. En les utilisant, il est possible de relier les contenus d'un musée ou d'une bibliothèque à ces ensembles existants et ainsi de créer du lien entre les institutions, chacune enrichissant ses propres données avec les données mises à disposition par les autres.

Premières réalisations emblématiques

Les avantages du web des données ont rapidement été compris par les institutions de la culture et du patrimoine, qui ont su s'en emparer pour réaliser des projets dont les plus emblématiques ont été salués pour leur degré d'innovation.

Bibliothèque Nationale de France

data.bnf.fr, le portail de la Bibliothèque Nationale de France, utilise les outils du web sémantique et s'inscrit dans une démarche d'ouverture des données.

Dans le cadre du schéma numérique de la Bibliothèque nationale de France, qui a pour but de rendre ses données plus visibles et plus utiles sur le web, ce portail a été mis en ligne dans une première version en 2011.

Il comporte désormais plus de 200 000 pages sur des auteurs, des œuvres et des thèmes, qui regroupent plus de 2 millions de références des catalogues de la BnF et plus de 140 000 liens vers Gallica. Cette belle réussite a reçu en 2013 le Stanford Prize for Innovation in Research Libraries et le Grand Prix Data Intelligence Awards.

Centre Pompidou

La numérisation des collections du Centre Pompidou, qui a apporté plus de 75 000 reproductions d'œuvres, a été la première étape du projet.

.../

La seconde phase fut marquée par l'ouverture en octobre 2012 du Centre Pompidou virtuel qui offre déjà la possibilité d'accéder à 95.000 ressources numériques, tous supports confondus. L'approche du web sémantique permet de lier les expositions et événements à toutes les œuvres concernées.

Les reproductions des 75.000 œuvres du Musée seront à terme accessibles. Une partie est d'ores et déjà en ligne, avec possibilité de zoomer en haute définition sur l'œuvre. Progressivement, des captations de performances, de conférences, interviews d'artistes et de commissaires d'exposition, documents écrits divers, catalogues d'expositions seront rendus disponibles.

D'autres projets « culture et patrimoine » tirent avantage du web sémantique

Le Musée des civilisations de l'Europe et de la Méditerranée ou MuCEM, inauguré en juin 2013, est dédié à la conservation, l'étude, la présentation et la médiation d'un patrimoine anthropologique relatif à l'aire européenne et méditerranéenne.

Un moteur de recherche sémantique facilite l'accès à l'information sur 900.000 objets, estampes, dessins, affiches, peintures, photographies, cartes postales, ainsi que 100.000 ouvrages composant sa bibliothèque et environ 1000 documents d'archives.

Les technologies du web sémantique ont été mises en œuvre pour enrichir les données avec le référentiel géographique Geonames, avec un référentiel temporel et avec un référentiel thématique : il est ainsi très simple de rebondir, depuis la fiche d'un objet, vers les objets provenant de la même région, vers les objets de la même période historique ou vers les objets de même nature.

L'Institut français est l'opérateur de l'action culturelle extérieure de la France. Sous la tutelle du ministère des Affaires étrangères, il contribue au rayonnement de la France à l'étranger dans un dialogue renforcé avec les cultures étrangères et répondre à la demande de France dans une démarche d'écoute, de partenariat et d'ouverture.

L'Institut français a créé le portail IFVerso, qui propose toute l'information sur les œuvres françaises traduites en langues étrangères : il facilite l'accès à une base de données en ligne de plus de 70 000 œuvres traduites vers plus de 40 langues, avec plus de 20000 auteurs, 20000 traducteurs et 8000 éditeurs du monde entier.

Le mouvement est lancé

Le web sémantique est en marche et le secteur de la culture et du patrimoine a su assez tôt en tirer avantage. Cela a notamment été rendu possible par des technologies et une expertise pour laquelle la France compte plusieurs acteurs à la pointe de l'innovation.

Aujourd'hui sont disponibles des solutions logicielles qui facilitent la mise en relation des données selon les principes et standards du web sémantique et favorisent l'accès à une information riche pour le grand public au travers de moteurs de recherche aussi puissants que simples d'emploi.