

Sujet de la thèse : Etude de l'apport et intégration des technologies du web sémantique au CMS WMaker et à son écosystème.

Entités de rattachement :

UMR CNRS 6134 – SPE et INIRIA Nice Sophia Antipolis

Contact :

Christophe Paoli

+33 (0) 4 95 45 02 09

christophe.paoli@univ-corse.fr

<http://www.christophepaoli.com>

Présentation succincte et explicite du projet

Abstract : Ce projet de thèse vise l'étude de l'apport et l'intégration des technologies du web sémantique au CMS WMaker et à son écosystème en deux axes : Mise à niveau « sémantique » du CMS WMaker et sémantisation du service de Web TV.

Présentation : Introduit par Tim Berners-Lee en 1998, le web sémantique ou web des données désigne un ensemble de technologies visant à rendre le contenu des ressources du web accessible et utilisable par les programmes et agents logiciels, grâce à un système de métadonnées formelles, utilisant notamment la famille de langages développés par le W3C, organisme de standardisation du web. Le Web sémantique est entièrement fondé sur le Web actuel dont il utilise les standards et les complète de quatre standards propres, ouverts, formant l'ossature du Web sémantique : RDF (modèle conceptuel permettant de décrire toute donnée), RDF Schema (langage permettant de créer des vocabulaires, ensembles de termes utilisés pour décrire des choses), OWL (langage permettant de créer des ontologies, vocabulaires plus complexes servant de support aux traitements logiques comme les inférences), SPARQL (langage de requêtes pour obtenir des informations à partir de graphes RDF). Conscient de la nécessité de franchir ce saut technologique, Wmaker souhaite qu'une étude très précise soit menée sur l'apport des technologies du web sémantique à leur solution de CMS et à leur écosystème. Scindé en un front office et un back office, le CMS de Wmaker répond à l'ensemble des critères et fonctionnalités définissant tout outil de gestion de contenu dit « web 2.0 » :

- La gestion du contenu textuel est assurée à travers un éditeur WYSIWYG comprenant les fonctionnalités suivantes : équipe de rédaction avec workflow de validation, classement des articles par rubriques et sous rubriques, restriction d'accès, accès gratuit/payant, publication différée, moteur de recherche avancée coté back-office, commentaires et trackbacks, etc.
- Au contenu textuel peut s'ajouter du contenu photo, vidéo et autres formats de fichier qui peut être géré au travers d'interface plus spécifiques : galerie, boutique, petites annonces, agenda et web TV.

- Au niveau du design, celui-ci est évolutif et est géré de manière totalement indépendante du contenu avec les fonctionnalités suivantes : proposition de maquettes, matrices de templates, affichage par modules avec drag & drop, design pour mobile, etc. Un ensemble de techniques visant à accroître la visibilité du contenu dans les moteurs de recherche est également offert.
- On retrouve également divers fonctionnalités telles que : la diffusion de newsletter, la création et la diffusion de podcasts, la gestion des statistiques, la possibilité de géolocalisation et de tagging/annotation, la possibilité de réseautage et de connectivité avec d'autres applications web 2.0.

De ces fonctionnalités et de l'adhésion d'une communauté forte a découlé une masse d'information importante à gérer avec près de 6.650 d'auteurs, 1.735.704 d'articles, 815.906 photos dans les galeries, 5.245.615 inscrits aux newsletters et 126.241 podcasts et videos. Afin de répondre au défi de cette « infobésité » les technologies du web sémantique semblent prometteuses. Plusieurs apports possibles ont été identifiés selon deux axes :

- Axe1 : Mise à niveau du CMS Wmaker. Ce premier axe propose de produire un état de l'art exhaustif des ontologies existantes en rapport avec les fonctionnalités du CMS WMaker afin de mettre à niveau celui-ci par rapport aux technologies du web sémantique. Les travaux suivants sont attendus :
 - Les technologies du web sémantique sont de plus en plus intégrées par les moteurs de recherche. Les récentes annonces concernant la prise en compte prioritaire du format RDFa par les moteurs de recherche impliquent une intégration des ontologies les plus populaires au sein du contenu produit par le CMS WMaker. Sont visées plus particulièrement le Dublin Core, FOAF, SIOC, GoodRelations et MOAT.
 - En s'inspirant de l'état de l'art sur les wiki sémantiques et en prenant en compte la nature multimédia du CMS, la question de l'intégration la génération de métadonnées à des interactions dans des interfaces WYSIWYG sera abordée.
 - Face à l'infobésité auquel peut être confronté tout webmaster, le mapping à des ontologies du contenu existant et/ou créé doit permettre une clarification des données et la mise à disposition d'un moteur de recherche sémantique. Celui-ci sera mis à disposition au sein du front et du back-office.
 - De l'aboutissement des deux travaux précédents pourra découler une refonte du back office de l'application. En effet la clarification des données et le mapping vers des ontologies permettra de mettre en évidence des relations de données jusqu'alors non explicites. Nous imaginons ainsi offrir de nouvelles fonctionnalités au sein de l'activité réseautage du CMS comme la recherche et la proposition de partenariats aux utilisateurs.
 - Enfin nous pensons que cette mise à niveau du CMS permettra un gain de productivité conséquent au sein de l'entreprise. Cette mise à plat du système d'information de WMaker permettra des échanges plus productifs entre les développeurs et une intégration plus rapide de nouveaux collaborateurs.

- Axe 2 : Web TV et web sémantique. Ce second axe plus novateur propose d'étudier l'apport des technologies du web sémantique concernant un nouveau service développé par le CMS de Wmaker appelé web TV. Plus précisément il s'agit de sites Internet dont les contenus sont à dominante vidéo. Nous proposons suite à un état de l'art dédié d'apporter des réponses en termes de fonctionnalités aux problématiques suivantes :
 - Annotation sémantique des contenus audio-visuels et de leurs fragments
 - Exploitation des annotations sémantiques dans la recherche et la composition d'éléments multimédia
 - Création de Web TV personnalisées en réponse à des requêtes d'utilisateurs en reposant sur un moteur de recherche sémantique.
 - Conception d'ontologies pour la mise en correspondance des contenus indexés et des demandes des auditeurs
 - Utilisation en mode collaboratif des fonctionnalités précédentes

Enjeux de la thèse

Abstract : Ce projet permettra une ouverture de l'université de Corse aux thématiques de recherche liées au web sémantique et aux travaux du W3C. De plus, il permettra un rapprochement plus fort encore de notre université de son environnement social et économique.

Présentation : La mise en place progressive du web des données engendra inévitablement une révolution des usages. Face à celle-ci, les logiciels qui permettront aux internautes de naviguer sur le web des données restent encore à peaufiner et à inventer. Outre le franchissement de ce saut technologique, plusieurs perspectives de nouveaux marchés se verront offerts à WMaker, notamment dans le monde des applications mobiles, du e-commerce et du e-tourisme. Mais au-delà de ces aspects purement économiques, des perspectives de recherche sont bien présentes, notamment autour des problématiques de représentation de la connaissance et d'interface homme-machine. Ainsi nous pensons que ce type de sujet de doctorat permettra une ouverture de l'université de Corse aux thématiques de recherche liées au web sémantique et aux travaux du W3C, organisme international de standardisation du web. Aussi ces travaux devraient permettre de rapprocher un peu plus encore l'université de Corse, et une partie de la recherche qui y est menée, de son environnement social et économique. Ce rapprochement favoriserait ainsi la création potentielle de richesse et d'emplois mais offrirait également des perspectives de carrières aux étudiants de l'université.