
Editorial

Le web 2.0 accompagné des technologies de l'interopérabilité a rendu la toile plus accessible et accéléré le développement de nouvelles applications riches tant par la nature des médias que par les fonctionnalités. Le document numérique est enrichi à la lumière des avancées technologiques, et peut être composé à partir de contenus distribués sur la toile. Les dimensions multilingues, multimédias et ontologiques peuvent plus facilement être appréhendées à travers les modèles et les architectures en vue de construire de nouvelles bibliothèques digitales. La mise à disposition de l'utilisateur de nouvelles possibilités pour construire et accéder à des contenus numériques répartis contribue à construire de nouveaux systèmes dont le but est entre autres de favoriser l'acquisition de connaissances par un plus grand nombre de citoyens.

Le web 2.0 va très probablement faire émerger de nouveaux usages. Déjà, on constate que les blogs et les wiki par exemple occupent les cyber-espaces et sont perçus comme de nouvelles possibilités offertes pour améliorer l'échange collaboratif entre internautes. Toutefois, si les potentialités du web 2.0 semblent ouvrir une nouvelle ère, nous n'en sommes qu'aux balbutiements et certainement à un stade « artisanal ». On est en droit de penser qu'une évolution plus structurée va permettre de professionnaliser les usages et l'accessibilité. Dans ce contexte, les contenus, au cœur des ressources du web, nécessitent de nouveaux outils pour prendre en compte les nouvelles technologies du web. Ces outils doivent être envisagés aussi bien sur le plan théorique en revisitant les modèles actuels de la recherche d'information, de l'indexation et de l'accès aux contenus, que sur le plan technique en proposant de nouvelles architectures et de nouvelles fonctionnalités qui tireraient profit des technologies telles que les services web, les techniques de balisage collaboratif, les techniques de syndication pour ne citer que celles-ci. Le web 2.0 doit également modifier en profondeur les usages car la participation à la production des contenus se fait de façon collaborative et nécessite alors de nouveaux outils qui apportent une valeur ajoutée dans la chaîne de production de contenus : création, gestion, accès, visualisation.

Ce numéro de document numérique tente à travers des articles très diversifiés de montrer l'étendue de la problématique et l'élargissement du document numérique à travers cette notion de contenu que l'on retrouve dans de nombreux champs d'investigation qui vont des bases de données aux bibliothèques digitales.

Nous avons décidé de retenir 7 articles que nous organisons en deux parties. La première partie regroupe les articles dans l'esprit habituel de « document numérique » où la dimension « document » voire « recherche d'information » est bien présente. La seconde partie se veut une ouverture sur des dimensions plus « sciences de l'information » pour approcher la problématique du contenu dans le web 2.0

autrement que par la simple transformation des représentations ou modèles de documents mais davantage à travers les usages et l'interprétation de l'information

Partie 1

L'article de Houda Oufaida et Omar Nouali est un état de l'art des approches de filtrage de documents. Les auteurs proposent une synthèse des différentes approches existantes en les classifiant en 3 catégories : systèmes à base de contenu, systèmes à base de filtrage collaboratif, et systèmes de filtrage hybrides. Le filtrage collaboratif est détaillé pour intégrer tous les travaux visant à utiliser les apports du web 2.0.

Ali Harb, Michel Plantié, Mathieu Roche, Gérard Dray, François Troussel et Pascal Poncelet présentent une nouvelle approche de création automatique de dictionnaires d'adjectifs intégrant la connaissance du domaine et en démontrent la validité et la pertinence par une expérimentation consistante menée sur des données réelles (comparativement à une méthode plus classique par apprentissage).

Dans l'article de Arnaud Blouin, Olivier Beaudoux et Stéphane Loiseau, différentes approches permettant la manipulation des données du web selon trois groupes : les approches basées sur les instances, celles basées sur les schémas, et celles dirigées par les modèles, sont explorées. Ce travail est réalisé dans le but d'effectuer une classification des approches pouvant être appliquées au problème de la manipulation des données du web.

Rachid Ahmed-Ouamer, Arezki Hammache font une proposition pour l'amélioration de la pertinence d'une recherche d'information en s'appuyant sur une Ontologie de domaine pour le e-learning et instanciée pour l'enseignement à distance de l'informatique. Dans ce contexte, l'ontologie est utilisée pour indexer des ressources pédagogiques par les concepts qui y sont définis. L'approche exploite également différents types de liens sémantiques (cause effet, implication, similaire, etc.) définis entre documents et/ou entre fragments de documents.

Partie 2

Dans ce cinquième article, Claire Fautsch, et Jacques Savoy s'intéressent à la problématique de la recherche d'information dans le contexte des blogs. Ils comparent l'utilisation de cinq techniques classiques sur des données de ce type extraites d'un corpus de TREC. Les problèmes liés à la « racinisation » sont aussi abordés ainsi que l'impact de la longueur des requêtes et de leur extension. L'article aborde une thématique très intéressante et au cœur du web 2.0 puisque le blog est un nouveau support pour véhiculer toute sorte d'informations au sens très large.

Luc Quoniam et Charles-Victor Boutet explorent diverses techniques employées pour diffuser l'information à travers le web (web 2.0). Ils nous sensibilisent aux principes mêmes du web 2.0 et montrent comment le peer-to-peer est un précurseur de la connectique informationnelle dans le monde du web 2.0.

Etienne Candel s'intéresse au social bookmarking en tant que pratique récente sur le web francophone. Dans sa tentative d'expliquer ce phénomène, Etienne Candel a considéré dans le cadre d'une expérimentation un corpus d'étude composé de quatre sites significatifs pour l'observation d'une pratique éditoriale dans un contexte culturel francophone : Blogmarks.net, Coomcoom.com, Bookmarks.fr et Bookeet.com.

Ce numéro thématique a offert l'opportunité pour des chercheurs d'évoquer l'accès aux ressources sur le web. Ces ressources pouvant être des contenus de toutes sortes, des documents, des sites ou encore des blogs. Ce dernier type de représentation de l'information sera très probablement appelé à se développer pour enrichir les représentations plus traditionnelles comme celle du document numérique. La revue « document numérique » est l'espace qui convient le mieux pour que la communauté scientifique puisse présenter les résultats de leurs investigations, qu'elles soient théoriques ou expérimentales, dans le domaine de l'accès aux ressources numériques du web 2.0.

Youssef Amghar
INSA de Lyon

