



Webmarks : Le marquage d'intérêt sur le Web de données

Nicolas Delaforge, Fabien Gandon

► To cite this version:

Nicolas Delaforge, Fabien Gandon. Webmarks : Le marquage d'intérêt sur le Web de données. EGC 2012, Jan 2012, Bordeaux, France. hal-00707499

HAL Id: hal-00707499

<https://hal.archives-ouvertes.fr/hal-00707499>

Submitted on 12 Jun 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Webmarks : Le marquage d'intérêt sur le Web de données

Nicolas Delaforge*, Fabien Gandon*

*INRIA Sophia Antipolis - Méditerranée
2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex
{prenom.nom}@inria.fr,
<http://www.inria.fr/>

Résumé. Depuis son apparition au sein du W3C, la définition de la ressource Web n'a cessé d'évoluer au delà du simple document. Lieu, service, concept d'ontologie, représentation d'un objet réel ou non, la ressource web est complexe et il nous a semblé que les outils à disposition des internautes pour sa manipulation, comme les bookmarks par exemple, n'exploitaient pas pleinement ces nouvelles dimensions. Dans cet article, nous présenterons le modèle Webmarks qui permet de préciser l'objet du marquage, la ressource, mais également l'intérêt de l'auteur de la marque. L'implémentation de ce modèle au sein du projet ISICIL sera également présentée et nous discuterons de son apport en comparaison des technologies existantes.

Mots-clés : Webmarks, Web sémantique, marquage d'intérêt

1 Introduction

Parmi toutes les possibilités que Tim Berners Lee avait ouvertes en inventant le Web, l'application la plus naturelle de cette technologie, la plus en accord avec les besoins de l'époque, fût l'application documentaire. C'est dans la continuité de cette logique que SGML, le standard documentaire de l'époque, servit de base à la rédaction de la première version de HTML et que les techniques de la documentation et de l'indexation servirent de guide pour le développement des premiers outils de navigation et de recherche online. Dans ce contexte de convergence numérique, certaines mnémo-techniques ancrées de longue date dans nos pratiques documentaires furent transposées telles quelles sur le Web ou dans les navigateurs sans que personne n'ait réellement le temps de réfléchir à l'impact que la dématérialisation du support pouvait avoir sur la signification de ces technologies.

Dans cet article nous nous intéresserons en particulier à l'une de ces mnémo-techniques, le *bookmark*, que certains ont préféré re-qualifier en *favori* ou *signet*¹. Nous verrons en quoi cette technologie s'est progressivement désynchronisée de la réalité du support Web, à cause notamment du changement de statut de la référence. Pour comprendre la fragilité de ces mécanismes référentiels inhérents aux bookmarks, il est utile de retourner quinze ans en arrière, dans les

1. Nous garderons ici le terme "Bookmark" afin de témoigner de l'origine "papier" de cette notion.

Webmarks : Marquer le Web de données

spécifications techniques qui ont fait évoluer le Web jusqu'à aujourd'hui. Dans une première section nous poserons donc les bases d'une théorie du support Web en retraçant brièvement les changements importants dans la définition de la ressource Web, puis dans les sections suivantes nous définirons ce qui constitue selon nous un marquage d'intérêt, nous ferons au passage une critique des systèmes de marquage existants et nous proposerons le modèle *Webmarks* en réponse aux lacunes identifiées. En particulier, nous verrons que les *Webmarks* permettent le marquage de ressources et pas seulement de documents. Enfin, nous terminerons par la présentation d'une implémentation de ce modèle au sein du projet de recherche ISICIL dans le cadre des activités de veille en entreprise.

2 Historique du marquage Web

Les premiers systèmes de stockage de références Web remontent au début des années 90, avec les premiers navigateurs. *Mosaic* en 1993 propose un système de *Hotlists* qui permet de stocker des références. Les premières versions d'*Opera*, *ViolaWWW* et *Cello* comportaient également ce genre de fonctionnalités. Le partage de listes de bookmarks apparaît assez tôt sur le Web avec *itList*[9] en 1996. C'est le début du *Social Bookmarking*[8]. Certaines plateformes connaissent un grand succès avec l'évolution du Web 2.0, c'est le cas de *Delicious*[6] et de *Connotea*[12]. Récemment, de nouveaux outils ont permis de partager des ensembles de bookmarks structurés sous forme de *mind maps* comme *Pearltrees* ou sous forme de parcours temporel comme *Jog The Web*. Le *Social Bookmarking* est indissociable du *Social Tagging* et les folksonomies servent de classifications légères pour naviguer dans les corpus de références Web. Néanmoins, du côté du Web Sémantique, en 2003, le navigateur Amaya a adopté une approche différente et s'est doté d'un système de bookmarks en RDF, issu du projet Annotea, qui permet notamment de lier la référence Web à des concepts de thésaurus pour une classification beaucoup plus rigoureuse des références[11] [10].

En 2004, avec l'expansion de flux RSS, Firefox intègre la fonctionnalité de *Live Bookmarks* dans son navigateur, ce qui permet de marquer des flux et de visualiser directement dans le navigateur les mises à jour. Avec les réseaux sociaux, le marquage Web a considérablement évolué. Les métadonnées présentes dans les pages économisent aux utilisateurs la tâche de renseigner un formulaire de marquage. Un simple bouton "I Like"[14] ou "+1" dans une page suffit pour ajouter une référence typée et documentée sur le profil Facebook ou Google+ de l'internaute. Le marquage et le partage sont instantanés et la marque est intégrée dans le flux de toutes les autres activités des utilisateurs.

Depuis Firefox, certains navigateurs proposent aux utilisateurs de développer des plug-ins qui permettent d'augmenter les capacités natives du logiciel. Grâce à cette approche participative, un florilège d'extensions a pu voir le jour et notamment certains outils permettant la création d'archives personnelles ou partagées, de tout ou parties de pages Web. Ces technologies communément appelées *Webscraps* connaissent aujourd'hui un développement conséquent. Parmi eux, nous citerons ScrapBook, Wozaik, Zotero, Diigo.

2.1 Évolutions du support

La nature du Web réside essentiellement dans son architecture technique. Celle-ci se compose de trois éléments clés : *URI*, *HTTP* et *Ressource*. Le marquage Web étant au carrefour de

ces trois notions, il est particulièrement sensible aux évolutions de leurs définitions. Or, parmi ces trois concepts, la ressource est la seule qui n'ait pas fait l'objet d'une définition claire sous forme de RFC [13]. Au contraire, sa définition est diluée au fil de la documentation proposée par le TAG².

2.1.1 Définition de la ressource Web : le serpent de mer du W3C

Les premiers documents qui parlent à la ressource datent de 1994 et le font de manière implicite. La ressource est vaguement désignée comme "tout ce qui peut être identifié par une URI" [2] ou comme un objet "disponible"(available) "qui peut être atteint via le réseau" [5]. La première véritable définition intervient en 1998 [4]. Les ressources sont désignées comme pouvant être identifiées par une URI et non nécessairement accessibles par le réseau. En janvier 2005, la définition de la ressource Web est étendue à nouveau de manière à inclure les concepts d'ontologies [3].

2.1.2 La fin du Web documentaire

Parallèlement aux discussions du W3C, le Web documentaire de première génération, a laissé la place au Web social "dynamique", dit Web 2.0. Les premières CGI, puis le PHP, ont ouvert la voie vers un Web de processus. À une URL donnée ne correspond pas un et un seul contenu mais potentiellement une infinité. Dans l'architecture du Web, chaque invocation HTTP d'une ressource par son URI génère potentiellement une nouvelle version.

En 2007, un projet apparaît parmi les W3C Task Force sous le nom de *Linking Open Data Project* (LOD³) et promeut l'idée d'un Web de données ouvert où les URI seraient déréférencées via le protocole HTTP. De cette manière, les relations entre les concepts pourraient se résoudre à travers le réseau, rendant ainsi accessible aux machines un graphe de données sémantiques et structurées à l'échelle du Web. Cette proposition de déréférencement des URI a suscité un vif débat au sein de la communauté Web sémantique, connu comme la "crise d'identité" du Web sémantique et a donné lieu à la rédaction d'un document connu sous le nom de "HTTP Range 14" [15] dans lequel le TAG propose un mécanisme de déréférencement des URI de ressources dites "non-informationnelles" sur le Web de données.

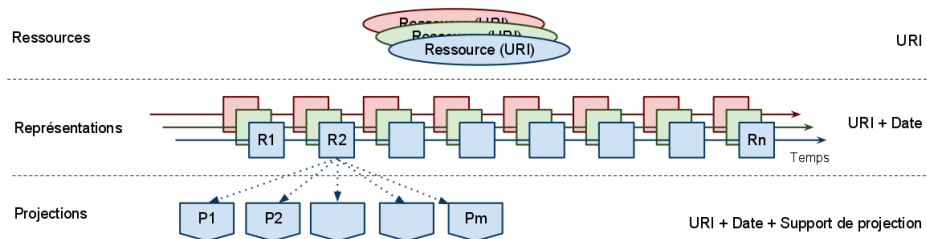


FIG. 1 – Couches d'abstraction du Web

2. Technical Architecture Group

3. <http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

Webmarks : Marquer le Web de données

Une ressource, identifiée par son URI, représente donc un calcul plutôt qu'un document et peut potentiellement générer une infinité de vues différentes qu'Halpin et Presutti [7] appellent *Web Representations*. La figure 1 illustre les couches d'abstraction identifiables dans l'architecture du Web qu'un utilisateur est susceptible de vouloir marquer. Dans ce contexte, quel est donc l'intérêt du Bookmark si les usagers du Web l'utilisent réellement, comme l'affirmait David Abrams en 1998 dans [1], comme une archive personnelle ?

3 Marquage d'intérêt

À défaut de référencer un contenu précis, le marquage d'une URI témoigne a minima d'un intérêt pour la ressource qui lui est associée. Le marquage peut prendre plusieurs formes, être accompagné ou non de métadonnées, néanmoins, la nature de cet intérêt reste souvent à définir. Il peut s'agir d'un intérêt pour la ressource de manière générale, informationnelle ou non, ou juste pour une représentation de la ressource. Cet intérêt peut encore être permanent ou ponctuel, lié à un événement, etc. Le Bookmark n'est pas le seul système de marquage d'intérêt pour un contenu en ligne. L'apparition des réseaux sociaux et du web participatif a donné lieu à de nombreuses technologies de ce genre, les outils de partage de références se sont multipliés. Nous allons dans cette section passer en revue, succinctement, les grandes fonctions qui se dégagent de cette offre logicielle.

Organiser : La plupart des outils disponibles pour l'annotation de ressources Web associent une description textuelle à la marque. Le tag représente la forme la plus réduite de métadonnée associées à une URL, a contrario du bookmark qui associe une description textuelle souvent plus riche : un titre, une description, etc. Ces métadonnées servent de base à l'organisation des références dans un corpus personnel. Les tags sont organisés en folksonomies, les bookmarks dans une hiérarchie de libellés souvent représentés comme des répertoires.

Partager : Le Web 2.0 a ajouté la dimension sociale aux technologies existantes. À travers des outils de *social tagging* et de *social bookmarking* comme Delicious⁴ ou Diigo⁵, les internautes ont la possibilité de stocker leurs annotations en ligne et de les partager.

Suivre : Bien que le succès des flux RSS semble avoir diminué depuis l'avènement de Twitter, les principes de la syndication restent les mêmes. L'abonnement à l'actualité d'une ressource témoigne d'un type d'intérêt à long terme. Ajouter un fil RSS dans un lecteur comme Google Reader ou s'abonner aux tweets d'un micro-blogger sont des activités proches du point de vue du type de marquage. La plupart des réseaux sociaux proposent également des "murs" pour suivre l'activité d'un membre du réseau. Les navigateurs disposent quasiment tous d'un type de marque-page dynamique qui permet d'accéder directement aux items du flux marqué.

Qualifier : Le bouton "*I Like*" de Facebook⁶ ou le "+1" de Google Plus⁷, constituent des mécanismes très simple de marquage d'intérêt pour une ressource, sans aucune métadonnée associée. L'apport principal de cette forme de marque d'intérêt réside essentiellement dans la puissance des plate-formes de réseaux sociaux couplées à ces dispositifs. Les marques sont automatiquement partagées aux contacts de l'intéressé permettant à ceux-ci d'évaluer la valeur de l'information marquée selon le nombre de personnes qui l'ont recommandée. Ces mécanismes

4. <http://del.icio.us/>

5. <http://www.diigo.com/>

6. <http://developers.facebook.com/docs/reference/plugins/like/>

7. <http://www.google.com/+1/button/>

de recommandation viennent renforcer les algorithmes complexes de ranking des moteurs de recherche en faisant de l'utilisateur un agent de qualification de l'information. Ces techniques de "social ranking" ont un autre avantage pour les éditeurs logiciels : Google et Facebook ont un modèle économique basé en grande partie sur la publicité. Les déclarations d'intérêt d'un utilisateur permettent de dresser un profil précis de celui-ci leur permettant de lui suggérer des publicités plus ciblées.

Préserver : Le Web scrapping est une réinterprétation moderne des outils de mirroring qui ont foisonné au début du Web. Ces technologies permettent aux utilisateurs d'extraire tout ou partie d'une représentation et de stocker le résultat (scrap) soit localement, comme le propose *ScrapBook* soit sur le cloud à la *Wozaik*. Certaines plate-formes de *social web scrapping*, comme *Zotero*, proposent de partager les *scrap*s à travers un réseau social et de les discuter, les commenter. L'enjeu principal de cette technologie réside essentiellement dans la stabilisation qu'apporte le processus d'archivage à un contenu. En effet, les *scrap*s sont systématiquement réintégrés dans un système de référence stable ce qui permet à l'activité référentielle d'avoir lieu sans risquer que la ressource change et que le contenu ne disparaisse. Contrairement aux autres systèmes exposés précédemment, le web scrapping témoigne d'un intérêt pour une représentation plutôt que pour la ressource.

4 Webmarks

4.1 Fonctionnalités

Préservation du contenu : L'ambition du modèle Webmarks est d'assister l'utilisateur dans la définition de l'objet de sa marque d'intérêt et dans la définition de l'intérêt qu'il porte pour l'URI marquée. La première étape est de préciser si l'utilisateur s'intéresse à la ressource dans sa globalité ou seulement au contenu instantané. Dans le second cas, le dispositif Webmark devra assurer la préservation du contenu à la manière d'un *Webscrap*. Les métadonnées associées au scrap seront sensiblement identiques à celles d'un Bookmark actuel, mais la référence marquée sera préservée et son URI stabilisée⁸.

Description par facettes : Si l'utilisateur semble s'intéresser plutôt à la ressource, le dispositif Webmark lui proposera d'attacher un certain nombre de facettes à la marque initiale :

`wm:GeoMark` : marque une ressource spatialement située, cette facette permet d'associer des coordonnées de géolocalisation à une ressource.

`wm:TemporalMark` : marque une ressource temporelle. Cette annotation permet d'ajouter à minima une date de début et une date de fin, l'utilisateur peut rajouter des événements supplémentaires de type `wm:Event` (voir schéma de l'ontologie).

`wm:LiveMark` : cette facette consiste à déclarer un intérêt durable pour une ressource à la manière des outils de syndication ou d'abonnement à l'actualité d'une ressource. L'utilisateur choisit une fréquence à laquelle le système ira analyser les éventuels changements de la ressource. Un système de notification avertira l'utilisateur que la ressource a changé en l'invitant à la consulter. D'autres services d'analyse différentielle ou d'archivage systématique pourraient être envisagés à partir de cette annotation.

8. Nous avons conscience que cette technique peut poser des problèmes du point de vue de la propriété intellectuelle et du droit d'auteur. Néanmoins, nous pensons que cette discussion n'a pas sa place ici.

Webmarks : Marquer le Web de données

`wm:SocialMark` : Pour les ressources qui sont associées à des individus, comme les pages de profil d'un réseau social, ou encore les pages personnelles, le système Webmarks proposera à l'utilisateur de renseigner quelques informations sur cette personne et sur les relations que l'utilisateur entretient avec elle. Cette fonctionnalité peut ajouter automatiquement la personne à la liste des contacts de l'utilisateur.

Parcours et marquage macroscopique : Le système de Webmarks doit permettre aux utilisateurs d'ajouter plusieurs ressources sous une même marque d'intérêt. Ces ressources peuvent être ordonnées de manière séquentielle, chronologiquement ou non. Nous désignons ce conteneur ordonné de référence `wm:Webtrack`.

Nous présentons ci-dessous deux scénarios représentatifs qui ont guidé la conception du modèle et la réalisation du prototype.

4.2 Scénarios

Scénario 1 : Contextualisation de la ressource Monsieur Green, reçoit un email concernant le CFP de la conférence EGC 2012. Il cherche donc "EGC 2012" dans Google et arrive sur la page de la conférence. Il estime que cette conférence l'intéresse et plus particulièrement les pages concernant le comité, le programme, les dates importantes et les informations pratiques. Il initialise un Webmark intitulé "EGC 2012", géolocalise la ressource à "Domaine du Haut Carré, rue du Haut Carré, 33400 Talence" et indique également que la ressource est temporalisée commence le 31 janvier 2012 et expire le 3 février 2012, il ajoute des dates supplémentaires : le 7 octobre 2011 qu'il intitule "Deadline résumé" et le 14 octobre en tant que "Deadline papier long". Ces dates s'inscriront automatiquement dans son calendrier.

Il décide d'ajouter une nouvelle ressource à sa marque EGC 2012. Pour cela, il convertit son Webmark en Webtrack (Webmark composite) et ajoute les ressources qui concernent le programme et les informations pratiques et demande à ce que les ressources soient "scrapées"⁹ sur son compte Webmarks.

Considérant que cette référence pourrait intéresser les autres membres de son équipe, il décide de leur partager cette marque. Les membres de son équipe reçoivent alors une notification comme quoi leur collègue Monsieur Green leur a partagé un Webmark. Ils peuvent alors consulter le descriptif de cette marque, voir le scrap et décider ou non de l'importer dans leurs corpus individuels. Lorsque la date d'expiration du Webmark sera arrivée à échéance, le statut de celui-ci changera alors automatiquement en "obsolète", permettant à Monsieur Green une gestion plus fine de ses références.

Scénario 2 : Marquage d'intérêt sémantiquement augmenté Monsieur Braun navigue sur le Web à la recherche d'information sur les voitures hybrides. Il arrive sur le site du constructeur Toyota qui vante les mérites de son modèle Prius. Sur la page des spécifications techniques de la Prius http://www.toyota.fr/cars/new_cars/prius/specs.aspx, il décide de poser une marque-web. L'extension Webmarks a analysé le contenu de la page Toyota, préalablement enrichie en RDFa et apprend donc que la page appartient au site de la marque Toyota, que c'est une page de spécifications techniques de la voiture Prius (dont il détecte

9. Les ressources sont extraites et archivées sur un serveur de *scrap*s, elles sont également accessibles de manière pérenne.

l'URI qui pointe vers le catalogue RDF Open Data du constructeur Nippon) et qu'elle contient les tags "Véhicule hybride, pile à combustible, Toyota". L'interface de marquage propose à Monsieur Braun différentes options afin de préciser l'intention de sa marque :

- Marque d'intérêt général pour l'entreprise Toyota
- Marque d'intérêt général pour le site web de l'entreprise Toyota
- Marque d'intérêt pour le modèle automobile "Prius"
- Marque d'intérêt pour la page des spécifications de la voiture "Prius"
- Marque d'intérêt pour l'un des tags "Véhicule hybride, pile à combustible ou Toyota"

Monsieur Braun indique son intérêt pour le modèle Prius, pour les véhicules hybrides et pour les spécifications de la Prius en particulier. Il demande à ce que le contenu de la page de spécifications soit archivé pour être sûr de le retrouver ultérieurement. Il choisit de rendre son Webmark public et de le taguer "Prius", "véhicule hybride" et "specs". Étant donné que Monsieur Braun a demandé l'archivage de la page, l'extension réalise un scrap de l'URL marquée et l'envoie au serveur de Webmarks pour son archivage. Lorsque Monsieur Braun souhaitera revenir sur sa marque, il lui sera possible de comparer les deux versions de la page, celle archivée et celle en cours. Il pourra réaliser une nouvelle capture et éventuellement demander à être notifié si des changements surviennent dans la page du site Toyota. Il demande également à être notifié des annotations concernant l'URI de la ressource "Prius".

5 Ontologie

L'ontologie Webmarks (figure 2) se base sur plusieurs standards du Web sémantique comme FOAF, Dublin Core et SIOC, ce qui permet aux annotations Webmarks d'être compatible avec les projets de recherche sur les réseaux sociaux sémantiques distribués.

Au cœur de l'ontologie se trouve le concept générique de Webmark qui agrège les propriétés communes à toutes les marques : titre, description, source. Ensuite, liées par une relation d'héritage, les différentes classes de marques viennent spécialiser le Webmark. Les possibilités d'héritage multiple en RDFS sont exploitées, ce qui fait qu'une marque-web peut être à la fois `wm:GeoMark` et `wm:LiveMark`, etc.

A l'interface du modèle, on voit que `wm:Webscrap` est une sous-classe de `isicil:Revision`. L'ontologie ISICIL n'est pas présentée dans cet article. Pour notre usage, cette classe permet à un `wm:Webscrap` de disposer d'un numéro de version et d'hériter par transitivité du type `sIOC:Item` ce qui permet à un `Webscrap` d'être intégré dans un `sIOC:Container` par la relation `sIOC:has_container`. La classe `wm:SocialMark` pointe vers un élément de type `rdfg:Graph`, soit un graphe nommé. Ce graphe doit contenir l'ensemble des informations sociales qui ont été saisies au moment du marquage. Cet élément est volontairement laissé ouvert de manière à pouvoir y agréger les informations issues de n'importe quel modèle (VCard, FOAF, Relationship, SIOC,...). Le `wm:Webtrack` est de type séquence RDF (`rdf:Seq`) de manière à ordonner les ressources de manière séquentielle.

L'ensemble des annotations d'un Webmark est contenu dans un graphe nommé de type `wm:MarkAction`. Ce graphe nommé nous permet d'associer un contexte de marquage aux informations qui constituent la marque en elle-même. `wm:MarkAction` étend `nicetag:TagAction` et par transitivité obtient le type `sIOC:Item`, ce qui lui permet de bénéficier de la propriété `sIOC:has_creator`. La figure 3 illustre un ensemble d'annotations RDF/XML

Webmarks : Marquer le Web de données

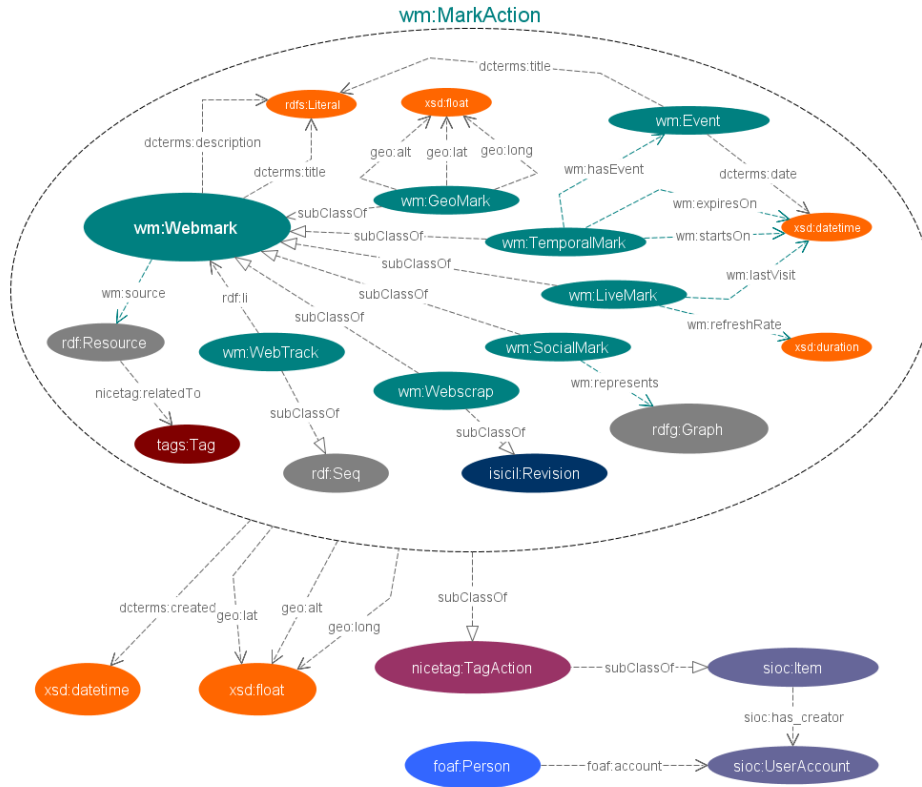


FIG. 2 – *Modèle Webmarks*

de Webmarks généré par le scénario numéro 1. Elle sont à la fois géolocalisées (ligne 9) et temporelles (ligne 15), contextualisées (ligne 30).

6 Implémentation du prototype au sein du projet ISICIL

Le prototype Webmarks a été réalisé au sein du framework SPREAN qui sert de base au développement dans le projet ISICIL.

ISICIL et SPREAN Framework

ISICIL est un projet ANR de trois ans qui vise à renouveler les pratiques de veille en entreprise, notamment par la mise en œuvre d'un ensemble d'outils issus du web 2.0 (réseaux sociaux, social tagging, social bookmarking, blogs, wiki, navigation augmentée, notifications, etc) basés sur les technologies du web sémantique. Ce projet a donné lieu à l'implémentation

```

1 <!--Webmark de base -->
2 <wm:Webmark rdf:about="http://ns.inria.fr/isicil/id/webmark/1"
3   cos:graph="http://ns.inria.fr/isicil/id/markaction/1">
4   <wm:source rdf:resource="http://egc2012.labri.fr/">
5   <dcterms:title>EGC 2012</dcterms:title>
6   <dcterms:description>Site de la conférence EGC 2012</dcterms:description>
7 </wm:Webmark>
8 <!--Informations géographiques -->
9 <wm:GeoMark rdf:about="http://ns.inria.fr/isicil/id/webmark/1">
10  <geo:alt>0</geo:alt>
11  <geo:lat>44.8117269</geo:lat>
12  <geo:long>-0.5972067000000152</geo:long>
13 </wm:GeoMark>
14 <!--Informations temporelles -->
15 <wm:TemporalMark rdf:about="http://ns.inria.fr/isicil/id/webmark/1">
16  <wm:startsOn rdf:datatype="&xsd;datetime">2012-01-31T00:00:00.000+0200</wm:startsOn>
17  <wm:expiresOn rdf:datatype="&xsd;datetime">2012-02-03T00:00:00.000+0200</wm:expiresOn>
18  <wm:hasEvent>
19    <wm:Event>
20      <dcterms:title>Deadline résumé</dcterms:title>
21      <dcterms:date rdf:datatype="&xsd;datetime">2011-10-07T23:59:00.000+0200</dcterms:date>
22    </wm:Event>
23    <wm:Event>
24      <dcterms:title>Deadline papier long</dcterms:title>
25      <dcterms:date rdf:datatype="&xsd;datetime">2011-10-14T23:59:00.000+0200</dcterms:date>
26    </wm:Event>
27  </wm:hasEvent>
28 </wm:TemporalMark>
29 <!--Informations contextuelles -->
30 <wm:MarkAction rdf:about="http://ns.inria.fr/isicil/id/markaction/1">
31  <dcterms:created rdf:datatype="&xsd;datetime">2011-09-21T01:55:20.361+0200</dcterms:created>
32  <sioc:has_creator rdf:resource="http://ns.inria.fr/isicil/id/account/mgreen"/>
33  <geo:alt>0</geo:alt>
34  <geo:lat>43.5804179999999</geo:lat>
35  <geo:long>7.12510199999997</geo:long>
36 </wm:MarkAction>

```

FIG. 3 – Annotation RDF/XML de Webmark

d'un framework complet destiné à améliorer les capacités de recherche et de collaboration des veilleurs sur leur poste de travail et en mobilité.

La plate-forme ISICIL comprend une API REST qui donne accès à :

- Un réseau socio-sémantique basé sur les technologies FOAF, SIOC et Relationship,
- Un système de tagging évolué compatible avec les technologies de thésaurus SKOS,
- Un système de description de ressources documentaires,
- Un modèle d'action et un système de suivi des activités individuelles,
- Un système de partage et de notifications,
- Un système de contrôle d'accès basé sur l'ontologie S4AC [16],
- Un endpoint SPARQL

Cette plate-forme est interfacée avec plusieurs clients : un wiki sémantique (SweetDeki), une application Android et une extension Firefox (ISICIL Browser).

C'est dans cette extension Firefox que le prototype Webmarks a été intégré. L'interface de Webmarking est couplée à l'interface de tagging. La complétion automatique sur les tags vient du serveur SPREAN. L'extension Firefox intègre le système de notification à travers le réseau socio-sémantique ISICIL. Ainsi les utilisateurs peuvent s'abonner à l'activité d'un autre

Webmarks : Marquer le Web de données

utilisateur ou encore à des thématiques (tags, concepts). La création d'un Webmark peut donc donner lieu à une notification si l'auteur du Webmark est lui-même suivi ou si les tags associés au Webmark font eux-mêmes l'objet d'un suivi.

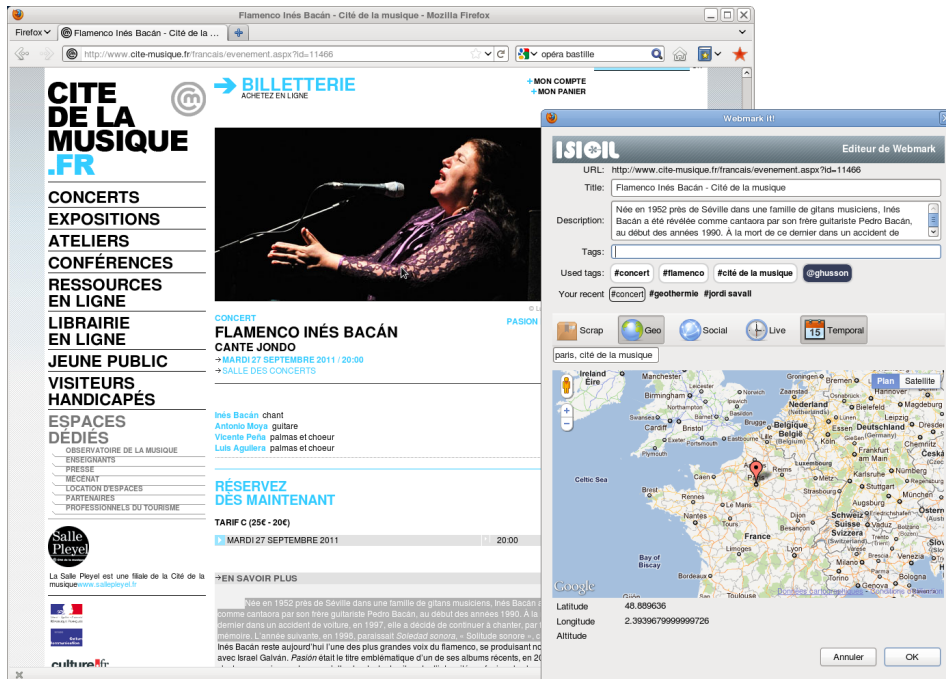


FIG. 4 – Interface Webmarks

La figure 4 montre le marquage d'une ressource de type concert à la cité de la musique. Le Webmark comporte des tags, il est à la fois `wm:GeoMark` et `wm:TemporalMark`.

Dans le projet ISICIL, la fonctionnalité de Webscrap consiste à créer une archive avec les contenus de la représentation et de téléverser l'archive en tant que pièce jointe d'une page du Wiki "SweetDeki". Le wiki propose un ensemble de fonctions de publication, manipulation et de réutilisation des scraps dans les pages de Wiki tout en offrant le stockage et la stabilisation nécessaire au bon fonctionnement des mécanismes de scrapping.

Évaluation du prototype

La campagne d'expérimentations ISICIL débute en janvier 2011. Elle s'effectuera en trois phases :

- un alpha-test de deux mois avec 4-5 utilisateurs membres du projet pour stabiliser la plate-forme ISICIL et corriger les interfaces,
- un bêta-test de deux mois avec 25 utilisateurs actifs aux profils divers. Leurs traces d'activité seront collectées et analysées.
- un test de deux mois à grande échelle parmi tous les membres du personnel Orange Labs.

Pour les tests, une plate-forme de suivi de bug et de soumissions de demandes d'évolutions fonctionnelles a été déployée sur le réseau expérimental de l'équipe. Le prototype Webmarks sera testé en tant que composant de production d'information et bénéficiera du wiki sémantique SweetDeki, des fonctionnalités de réseau social, de recommandation et de partage de la plate-forme ISICIL.

7 Conclusions

Dans cet article, nous avons proposé un modèle générique et extensible pour la représentation des Webmarks. Ces annotations ont pour vocation de se substituer aux bookmarks actuels dans les navigateurs et dont nous avons fait la critique dans la première partie de cet article. À nos yeux, l'apport du modèle Webmarks est multiple :

- il couvre la plupart des fonctionnalités identifiées dans la section 3,
- il permet de distinguer l'intérêt pour une ressource de l'intérêt pour une représentation et offre les services de préservation adéquats,
- il adapte son niveau de description de la ressources aux exigences de l'utilisateur. Que l'utilisateur veuille juste associer un titre à une URI ou au contraire s'il veut décrire en détail sa ressource, le modèle s'adapte à ses besoins grâce au système de facettes par héritage multiple.
- il exploite les standards du Web sémantique. Il est donc *web of data compliant*.

Perspectives La suite de ces travaux nous conduira à implémenter les fonctionnalités manquantes dans le prototype actuel (SocialMark et LiveMark).

Ensuite à plus long terme, nous envisageons d'étendre la déclaration d'intérêt à d'autres types de ressources (Web services, Concepts du Web de données, Applications).

D'autres chantiers pourraient être lancés en parallèle comme celui de l'analyse globale des champs d'intérêts individuels (l'ensemble des Webmarks d'un utilisateur) pour la suggestion de ressources à consulter, de personnes à ajouter à son réseau social, etc.

Remerciements L'ANR et le projet ISICIL (contrat ANR-08-CORD-011-05).

Références

- [1] David Abrams, Ron Baecker, and et al. Information archiving with bookmarks : Personal web space construction and organization. In *ACM SIGCHI '98*, pages 41–48. ACM Press, 1998.
- [2] Tim Berners-Lee. Rfc 1630 - universal resource identifiers in www, June 1994.
- [3] Tim Berners-Lee, R. Fielding, and L. Masinter. Rfc 3986 - uniform resource identifier (uri) : Generic syntax, January 2005.
- [4] Tim Berners-Lee, R. Fielding, L. Masinter, and U.C. Irvine. Rfc 2396 - uniform resource identifiers (uri) : Generic syntax.u.c. irvine, August 1998.
- [5] Tim Berners-Lee, L. Masinter, and M. McCahill. Rfc 1738 - uniform resource locators (url)university of minnesota, December 1994.

Webmarks : Marquer le Web de données

- [6] Matt Biddulph. Introducing del.icio.us, November 2004.
- [7] Harry Halpin and Valentina Presutti. An ontology of resources : Solving the identity crisis. In *ESWC 2009*, LNCS 5554, pages 521–534. Springer-Verlag, 2009.
- [8] Tony Hammond, Timo Hannay, Ben Lund, and Joanna Scott. Social bookmarking tools : A general review. *D-Lib Magazine*, 11(4), April 2005.
- [9] LaJean Humphries. itlist and other bookmark managers, January 2000.
- [10] Marja-Riitta Koivunen. Annotea and semantic web supported collaboration. In *Proceedings ESWC 2005, UserSWeb workshop*.
- [11] Marja-Riitta Koivunen, Ralph Swick, and Eric Prud'hommeaux. Annotea shared bookmarks. KCAP03 workshop, October 2003.
- [12] Ben Lund, Tony Hammond, Martin Flack, and Timo Hannay. Social bookmarking tools : A case study - connotea. *D-Lib Magazine*, 11(4), April 2005.
- [13] Alexandre Monnin. Qu'est-ce que le web sémantique ? Article de blog, Ministère de la Culture et de la Communication, <http://cblog.culture.fr/2011/09/07/web-semantique-iri-opendat>, 2011.
- [14] David Recordon. The open graph protocol : Understanding the design decisions, 2010.
- [15] W3C Technical Architecture Group (TAG). *Dereferencing HTTP URIs*. W3C, 2007.
- [16] Serena Villata, Nicolas Delaforge, Fabien Gandon, and Amélie Gyrard. An access control model for linked data, October 2011. 7th International IFIP Workshop on Semantic Web & Web Semantics (SWWS 2011), à paraître.

Summary

As defined by the W3C TAG, the web resource is one of the most important part of the web architecture. Since its apparition, the resource definition evolved through URI/URN/URL RFCs and RDF specifications. From document to service or application as well, the resource is now considered as abstract as a concept. Therefore a web resource can denote almost everything: real or virtual objects, it also can be set up in a spatio-temporal context, in a social network or in a giant open data conceptual graph. Then, it appeared that contemporary tools for resource manipulation, like bookmarks, are considerably outdated regarding these new dimensions of web resources. The french ANR research project ISICIL studies the impact of ICT on the business intelligence workflows and proposes to improve the performance of these processes through semantic based technical solutions. In this context, we developed a tool for marking and sharing web resources of interest. In this article we will present the Webmarks prototype and we will discuss about the underlying ontology and how it contributes to the philosophical engineering discussion.