Adnosco : gérez les données que vous diffusez !

Nadia Bennani, Emmanuel Gaude, Előd Egyed-Zsigmond, Philippe Lamarre

▶ **To cite this version:**

Nadia Bennani, Emmanuel Gaude, Előd Egyed-Zsigmond, Philippe Lamarre. Adnosco : gérez les données que vous diffusez !. Catherine Faron-Zucker. IC - 25èmes Journées francophones d'Ingénierie des Connaissances, May 2014, Clermont-Ferrand, France. Journées francophones d Ingénierie des Connaissances, pp.287-290, 2014. hal-01016020

HAL Id: hal-01016020
https://inria.hal.science/hal-01016020

Submitted on 27 Jun 2014

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.

Adnosco : gérez les données que vous diffusez !

Nadia Bennani, Emmanuel Gaude, Előd Egyed-Zsigmond, Philippe Lamarre

Université de Lyon
CNRS, INSA-Lyon, LIRIS, UMR5205, F-69621, France
firstname.lastname@liris.cnrs.fr

Résumé :
Le nombre de formulaires en ligne a explosé avec le développement du web. C'est le moyen le plus répandu pour acquérir des informations auprès des utilisateurs. Actuellement, ces données sont stockées par les industriels ou les administrations. Les utilisateurs ont un rôle entièrement passif qui les rend dépendants pour la gestion de leurs informations. Nous proposons de démontrer que, convenablement modélisées, la gestion de ces informations par les utilisateurs peut leur être bénéfique aux deux parties : utilisateurs et fournisseurs de services. Cette démonstration a donc pour vocation de présenter un modèle et un prototype basés sur la qualification sémantique de formulaires en ligne qui améliorent la productivité de l'utilisateur.
Mots-clés : Formulaire web, alignement sémantique, complétion.

1 Introduction

De nos jours, la plupart des sites Web archivent les informations personnelles de leurs utilisateurs. De plus, ils disposent de suffisamment d'outils (CRM, fouille de données, etc) pour permettre l'acquisition, le stockage, l'accès, la gestion et l'exploitation des informations obtenues auprès de l'utilisateur. A l'inverse, l'utilisateur final ne dispose d'aucun moyen pour effectuer les mêmes opérations sur ses propres données. Pour illustrer cette lacune, posons nous la question suivante : "quel système permet à un utilisateur de déterminer les informations qu'il a transmises, à quels sites, dans quel contexte ou objectif ?". Sans remettre en cause l'utilité et le droit qu'ont les sites marchands à stocker les données de l'utilisateur, il nous paraît naturel et très complémentaire de donner à ce dernier les moyens de gérer ses propres données. De plus, cela permet d'imaginer d'autres scénarios. Par exemple, suite à un achat en ligne, les informations acquises peuvent être transmises à d'autres applications : *montant* au gestionnaire de comptes bancaires, *date estimée de livraison* à l'agenda. . .. Actuellement, à notre connaissance, ce type d'usage n'existe pas.

Cette approche orientée utilisateur rejoint celle développée par Berkman Center For Internet & Society à l'Université d'Harvard qui propose la notion de *VRM* [3] (Vendor Relationship Management), complémentaire à celle plus connue des *CRM* [5] qui elle, est plus bénéfique aux marchands.

Pour œuvrer dans ce sens, dans cette démonstration, nous présentons *Adnosco*, un outil dédié à la gestion et à l'intégration des données utilisateur. *Adnosco* s'attaque aux problèmes d'acquisition, de stockage, de structuration de données, dans le but de les interroger et de les exploiter. Par manque de place, dans cet article, nous ne discuterons pas les problèmes d'acquisition. L'exploitation des données sera vue sous l'angle restreint de l'assistance à la complétion de formulaires Web, bien qu'en réalité, *Adnosco* permettrait l'exploitation des données utilisateurs dans un contexte plus large.

Une étude récente [2] démontre que la saisie de formulaires est la cause du renoncement à un achat en ligne pour 24% des utilisateurs. Les fonctionnalités d'auto-complétion et de pré

remplissage ont prouvé leur aptitude à augmenter la productivité des utilisateurs dans des environnements professionnels (outils de gestion de stock, de clients…). C'est ce qui a motivé notre choix de la complétion pour démontrer la capacité d'*Adnosco* à gérer et exploiter les données de l'utilisateur avec précision et efficacité. Les solutions actuelles permettent de compléter un champ d'un formulaire uniquement en proposant un ensemble de valeurs issues du même champ dans d'autres instances du même formulaire ou pour des informations très spécifiques (par exemple, le nom, l'adresse, les numéros de téléphone, les cartes de crédits…). La complétion dans *Adnosco* améliore considérablement ces fonctionnalités en proposant pour un champ, toutes les valeurs sémantiquement pertinentes, issues de tous les formulaires préalablement remplis. Nous proposons 2 outils de complétion, l'un syntaxique, l'autre sémantique.

2 Modèle de gestion de données

Adnosco est basé sur un modèle de stockage de données en trois parties (figure 1). La première, *data management*, analyse chaque formulaire, lors de sa soumission, pour acquérir les données transmises dans les champs de celui-ci (ou suite à une demande explicite de l'utilisateur) et stocke les valeurs ainsi obtenues dans une base de données en les reliant à ce formulaire (uri, identifiants, date, types et valeurs des champs). La seconde, *semantic concepts*, définit des *concepts*. Son rôle est de fournir les outils sémantiques, comme les schémas, ontologies, concepts, attributs, types, relations,… servant à caractériser les informations diffusées. La troisième partie, *semantic qualification*, relie les deux premières. Afin de rendre aisé et précis, l'alignement sémantique des formulaires web , nous introduisons la notion de *concept matérialisé* (CM). Un concept matérialisé formalise l'expression d'une occurrence de concept sémantique (par exemple une personne) dans un formulaire (voyageur) en établissant des correspondances entre les propriétés que le concept matérialisé (voyageur) hérite du concept (personne) et les champs du formulaire.

L'assistance syntaxique est relativement limitée car elle ne s'appuie que sur la première partie. L'assistance sémantique utilise les 3 parties et peut donc être bien plus précise sur l'usage des informations collectées.

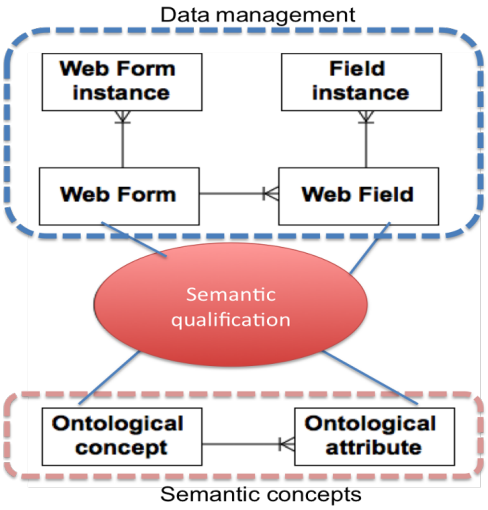

FIG. 1 – Modèle de stockage données d'*Adnosco*

C'est grâce aux concepts matérialisés que la complétion et le pré remplissage des champs sont traités correctement dans le cas où plusieurs occurrences d'un même concept sont présentes dans un même formulaire. Par exemple, un formulaire de voyage peut impliquer plusieurs personnes. Sa qualification sémantique fera apparaître autant de concepts matérialisés '*Personne*' distincts que nécessaire.

Lorsque l'utilisateur veut remplir un champ associé à un concept matérialisé (ex. Voyageur1), le système se concentre sur les champs de ce CM et propose pour complétion des valeurs correspondant aux propriétés d'instances du même concept (ex. Personne) compatibles avec les informations déjà saisies. L'aide au remplissage de formulaires web est donc précise et efficace. Une description plus détaillée d'un scénario illustrant les avantages d'*Adnosco* est décrite dans [1].

3 Conclusion et travaux futurs

Dans cette démonstration, nous proposons *Adnosco*, un système de gestion des données, centré utilisateur. *Adnosco* offre une solution d'acquisition, de stockage et d'interrogation des données transmises à des tiers via des formulaires web. Son service de complétion fournit une aide au remplissage des formulaires grâce à l'exploitation de données déjà transmises. Il met en œuvre des techniques syntaxiques et sémantiques (concepts matérialisés) qui s'appliquent à tout type d'information et qui permettent, par exemple, de traiter correctement des formulaires web contenant plusieurs instances d'un même concept. En résumé, *Adnosco* généralise des solutions comparables [2, 4] et améliore la productivité de l'utilisateur que ce soit dans le cadre d'activités personnelles ou professionnelles. *Adnosco* ne s'oppose pas aux activités de collecte d'informations de type CRM, mais il apporte aux utilisateurs une solution complémentaire rétablissant l'équilibre des rôles dans la gestion de leurs données personnelles. De leur côté, les organismes collecteurs peuvent guider la saisie de l'utilisateur en s'appuyant sur cet outil (qualification sémantique de leurs formulaires) pour améliorer la qualité des informations obtenues.

Références

[1] ADNOSCO (2014). adnosco. http://adnosco.liris.cnrs.fr/doku.php?id=scenario.

[2] DASHLANE (2013). Dashlane. https://www.dashlane.com/download/Dashlane_IFOP_release_2013-03-26_en.pdf.

[3] HAVARD (2007). Vendor management system. http://cyber.law.harvard.edu/.

[4] MIT (2013). openpds. http://openpds.media.mit.edu/.

[5] WIKIPEDIA (2013). Customer management system. http://en.wikipedia.org/wiki/Customer_relationship_management.