The Semantic Web: From Representation to Realization

Kristinn R. Thórisson[1,2], Nova Spivack[2], and James M. Wissner[2]

[1] Center for Analysis & Design of Intelligent Agents
and School of Computer Science, Reykjavik University
Menntavegur 1, 101 Reykjavik, Iceland
[2] Radar Networks, Inc.
410 Townsend Street, Suite 150, San Francisco, CA 94107 USA
thorisson@ru.is, {nova,jim}@radarnetworks.com

Abstract. A semantically-linked web of electronic information – the Semantic Web – promises numerous benefits including increased precision in automated information sorting, searching, organizing and summarizing. Realizing this requires significantly more reliable meta-information than is readily available today. It also requires a better way to represent information that supports unified management of diverse data and diverse Manipulation methods: from basic keywords to various types of artificial intelligence, to the highest level of intelligent manipulation – the human mind. How this is best done is far from obvious. Relying solely on hand-crafted annotation and ontologies, or solely on artificial intelligence techniques, seems less likely for success than a combination of the two. In this paper describe an integrated, complete solution to these challenges that has already been implemented and tested with hundreds of thousands of users. It is based on an ontological representational level we call *SemCards* that combines ontological rigour with flexible user interface constructs. SemCards are machine- and human-readable digital entities that allow non-experts to create and use semantic content, while empowering machines to better assist and participate in the process. SemCards enable users to easily create semantically-grounded data that in turn acts as examples for automation processes, creating a positive iterative feedback loop of metadata creation and refinement between user and machine. They provide a holistic solution to the Semantic Web, supporting powerful management of the full lifecycle of data, including its creation, retrieval, classification, sorting and sharing. We have implemented the SemCard technology on the semantic Web site *Twine.com*, showing that the technology is indeed versatile and scalable. Here we present the key ideas behind SemCards and describe the initial implementation of the technology.

Keywords: Semantic Web, Ontologies, Knowledge Management, User Interface, SemCards, Human-Machine Collaboration, Twine.com, Metadata.

N.T. Nguyen and R. Kowalczyk (Eds.): Transactions on CCI II, LNCS 6450, pp. 90–107, 2010.

1 Introduction

Intelligent automated retrieval, manipulation and presentation of information defines the speed of progress in much of today's high-technology work. In a world where information is at the center, any improvement is welcomed that can help automate even more of the massive amounts of data manipulation necessary. In many people's vision of the Semantic Web machines take center stage, based on a deeper knowledge of the data they manipulate than currently possible. To do so calls for metadata – data about the data. Making machines smarter at tasks such as retrieving relevant information at relevant times automatically from the vast collection, even on today's average laptop hard drive, requires much more meta-information than is available at present for this data.

Accurate metadata can only be derived from an understanding of content; classifying photographs according to what they depict, for example, is best done by a recognition of the entities in them, lighting conditions, weather, film stock, lens type used, etc. Authoring metadata for images by hand, to continue with this example, will be an impossible undertaking, even if we limited the metadata to surface phenomena such as the basic objects included in the picture, as the number of photographs generated and shared by people is increasing exponentially. Powertools designed for *manual metadata creation* would only improve the situation incrementally, not exponentially, as needed.

Although text analysis has come quite a long way and is much further advanced than image analysis, artificial intelligence techniques for analyzing text and images have a long way to go to reliably decipher the complex content of such data. The falling price of computing power could help in this respect, as image analysis is resource-intensive. This will not be sufficient, however, as *general-purpose* image analysis (read: software with "commmon sense") is needed to analyze and classify the full range of images produced by people based on content. On the one hand, achieving the full potential of a semantic web, leaving metadata creation to current AI technologies, will not be possible as these technologies are simply not powerful enough. This state of affairs may very possibly extend well beyond the next decade. On the other hand, because the growth of data available online is rising exponentially, and can be expected to continue to do so, manual metadata entry will never catch up to the extent necessary for significant effect. Creating the full set of ontologies by hand required for adequate machine manipulation would be a Herculean effort; waiting for the adequate machine intelligence could delay the Semantic Web for decades.

Does this mean the semantic web is unrealizable until machines become significantly smarter? Not necessarily. While we believe that neither hand-crafted ontologies nor current (or next wave) artificial intelligence techniques alone can achieve a giant leap towards the Semantic Web, a clever combination of the two could potentially achieve more than a notable improvement. The idea is that if online manual labor could somehow be augmented in such a way that it supported automatic classification, making up for its weak points, this could help move the total amount of semantically-tagged data closer to the 100% mark and help automatic processes get over the well-known "90% accuracy brick wall".

For us the question about how to achieve the vision of the Semantic Web has been, *What kind of collaborative framework will best address the building of the Semantic Web?* Most tools and methodologies designed for automating data handling are not suitable for human usage – the underlying data representations is designed for machines in ways that are not meant for human consumption. Data formats designed exclusively for human usage, such as e.g. HTML, are not suitable for machine manipulation – the data is unstructured, the process is slow, error prone and ultimately, to make it work, calls for massive amounts of machine intelligence that are well beyond today's reach.

This line of reasoning has resulted in our two-prong approach to the creation of the Semantic Web: First, we develop a system for helping people take a more structured approach to their data creation, management and manipulation and second, we develop automatic analysis mechanisms that use the human-provided structured data and framework to expand the semantic classification beyond what is possible to do by hand. We have already achieved significant progress on the first part of this approach; the second part is also well under way. Our method facilitates an iterative interaction loop between the user's information input, the automated extension of this work and subsequent monitoring of feedback on the extensions from the user.

Semantic Cards, or *SemCards*, is what we call the underlying representation of our approach. It is a technology that combines ontology creation, management/usage with the user interface in a way that supports simultaneously (a) human metadata creation, manipulation and consumption, (b) expert-user creation and maintenance of ontologies, and (c) automation services that are augmented by human-created meaningful examples of metadata and semantic relationship links, which greatly enhances their functionality and accuracy. SemCards provide an intermediate ontological representational level that allows end-users to create rich semantic networks for their information sphere.

One of the big problems with automation is low quality of results. While statistics may work reasonably in some cases as a solution to this, for any single individual the "average user" is all too often too different on too many dimensions for such an approach to be useful. The SemCard intermediate layer encourages users to create metadata and semantic links, which provides underlying automation with highly specific, user-motivated examples. The net effect is an increase in the possible collaboration between the user and the machine. Semi-intelligent processes can be usefully employed without requiring significant or immediate leaps in AI research.

From the users' perspective what we have developed is a network portal where they can organize their own information for personal use, publish any of that information to any group – be it "emails" addressed to a single individual or photo albums shared with the world – and manage the information shared with them from others, whether it is documents, books, music, etc. Under the hood are powerful ontology-driven technologies for organizing all categories of data, including access management, relational (semantic) links and display policies, in a way that is relatively transparent to the user. The result is a system that offers improved automation and control over access management, information organization and display features.

Here we describe the ideas behind our approach and give a short overivew of a use-case on the semantic Web site Twine.com. The paper is organized as follows: First we review related work, then we describe the technology underlying SemCards and explain how the are used. We then describe our Web portal Twine.com, where we have implemented various user interfaces for enabling the use of SemCards in a number of ways, including making semantically rich Web bookmarks, notes, blogs and semantically-annotated uploads.

2 Related Work

The full vision of the Semantic Web will require significant amounts of metadata, some of which describes entities themselves, other which describes relationships between entities. Two camps can be seen proposing rather different approaches to this problem. One extreme claims that manual creation of metadata will never work as it is not only slow and error-prone, the level to which it would have to be done would go well beyond the patience of any average user – quite possibly all. To this camp the only real option is automation. The other camp points out that automation is even more error-prone than manual creation, as current efforts to automatic semantic annotation on massive scales produces only moderate results of between 80% and 90% correct, at the very best [1]. They claim that the remaining 10% will always be beyond reach because it requires significant amounts of human-level intelligence to be done correctly. Further, as argued by Etzioni and Gribble [2], metadata augmentation has quite possibly not been done by the general user population because they have seen no benefits in doing so. Lastly, this camp points to the massive amounts of tagging and data entry done on sites such as Wikipedia, Myspace and Facebook as a proof of point that end-users are quite willing to provide (some amount of) metadata. Giving them the right tools might change this. Applications that connect casual end-users with ontologically-driven content and processes are, nevertheless, virtually non-existent.

Many efforts have focused on building digital content management with a focus on the object. Of these, our technology bears perhaps the greatest resemblance to the *Buckets* of Maly et al. [3] which are "self-contained, intelligent, and aggregative ... objects that are capable of enforcing their own terms and conditions, negotiating access, and displaying their contents". Like SemCards, Buckets are fairly self-contained, with specifications for how they should be displayed. Buckets grew out of Kahn and Wilensky's [4] proposed infrastructure for digital information services. Key to their proposal was the notion of digital object, composed of essentially the two familar parts, data and metadata. The subsequent work on FEDORA [5] saw the creation of an open-source software framework for the "storage, management, and dissemination of complex objects and the relationships among them" [6]. Buckets represent a focus on storing content in digital libraries, most likely manipulated by experts. In contrast, SemCards aim at enabling casual end-users to create metadata. Buckets are targeted to machime manipulation; SemCards are aimed at machine *manipulation* as well, but more importantly at supporting *automatically generated* meta-information. SemCards also differ from Buckets in that they are especially designed to be sharable between multiple users over mixed-architecture networks.

The Haystack [7] and Chandler[1] projects were efforts to create new user interfaces for wiewing and working with semantic objects. While this work was important – and in many ways still is – it also shows how difficult it is to lead such efforts to ultimate fruition while addressing all the key issues that must be solved. Our work on PersonalRadar followed a similar path[2], albeit always with the ultimate objective of solving the hard problems related to deployment over a WAN.

The separate representation layer provided by SemCards is a key difference between prior efforts and ours. They enable ontologically-driven constructs to be collaboratively built by ontology specialists, algorithms and end-users, encouraging them to provide examples to improve the automation. Because of this, SemCards are tolerant to end-user mistakes; the casual Internet user is not initiated to invest a lot of time in understanding the intricacies of the kinds of advanced ontologies required. Separating the two makes the automation systems more robust to manual input errors.

Other important differences between our approach and prior work are an integrated ability to share data between individuals and groups of users over a network, with complex policy control over access and sharing, and the flexible use of SemCards to represent metadata for real-world objects and hypothetical constructs - as "library index cards for digital content, physical things and abstract ideas".

Although current enterprise portals are capable of organizing group or team information, they are often inaccessible to the public or to individuals, and they are expensive as they are highly monolithic. Even less utilitarian and intelligent with respect to organizing information are the popular online search engines which are deisgned for largely unstructured data. Furthermore, these typically organize information and data by relevance to keywords. We have built a network portal, *Twine.com*, for deploying the SemCard technology. Twine.com provides a test of the strength of our semantic object framework when deployed over the Internet and working in an integrated, coordinated manner. Our work sets itself apart from prior work on the Semantic Web in that it has already been tested with a relatively large number of end users, with measurable results.

[1] `http://chandlerproject.org`

[2] PersonalRadar was a desktop application that we developed around the same time that Chandler became public, and in some ways it presented similar solutions to the semnatic interface; the semantic search/filtering interface for PersonalRadar was, however, vastly superior to anything we have seen so for proposed for that purpose. Unfortunately the numerous excellent interface ideas developed for PersonalRadar are still not supported by Twine as it is virtually impossible to implement these methods over a standard network link.

3 SemCards: Semantic Objects for Collaborative Ontology-Driven Information Management

A single *SemCard* can be characterized as an intuitive user interface construct that bridges between a user and an underlying ontology that affords all the benefits of a Semantic Web such as automatic relationship discovery, sorting, data mining, semantic search, etc. Together many SemCards form semantic nets that are in every way the embodiment of what many have envisioned the Semantic Web to be. Instead of being complex, convoluted and non-intuitive as any machine-manipulatable ontology will appear to the uninitiated (c.f. [8]), SemCards provide a powerful and intuitive interface to a unified framework for managing information.

As mentioned above, SemCards form an intermediate separation layer between ontologies and the user interface. By isolating the stochastic nature of end-user activity from underlying semantic networks built with ontological rigour, two important goals are achieved. First, end-users are encouraged to create metadata for their content, as the input methods are familiar and straight-forward. SemCards shield the deep ontology from being affected by end-user activity. This does not only help stabilize the system, it also helps the automation processes from having to deal with the "ground shifting from underneath". Second, the automation processes are provided with manually-created semantic nets, created directly and continuously by end-users, that serve as examples and can be used to improve the automatic metadata creation. The net result of this is a significant improvement in automation quality and speed, including automation of many tedious details of information management such as data sharing policy maintenance, indexing, sorting – in fact, the of the full data management lifecycle.

3.1 Structure of a SemCard

In its simplest version a SemCard will appear to the user as a form with fields or slots. A SemCard has one template and one or more instances, corresponding roughly to the object-oriented programming concepts of object template/class, and object instance, respectively. Under the hood their slots are ontologically defined; however, the end user normally does not see this. To take an example, a SemCard for holding an e-mail message may look exactly like any interface to a regular email program. However, the slots ("To:", "From:", etc.) reference an ontology that defines what kinds of data each slot can take, what type of information that is, etc. The e-mail SemCard, when created, will contain information about who authored which part of the content and when. Additionally, the author will not simply be a regular "From" but have a link to the SemCard representing the author of the email SemCard.

No executable code. An important feature of SemCards is that they are completely passive – they do not carry with them any executable code: We have entirely separated the services operating on the SemCards from the SemCards

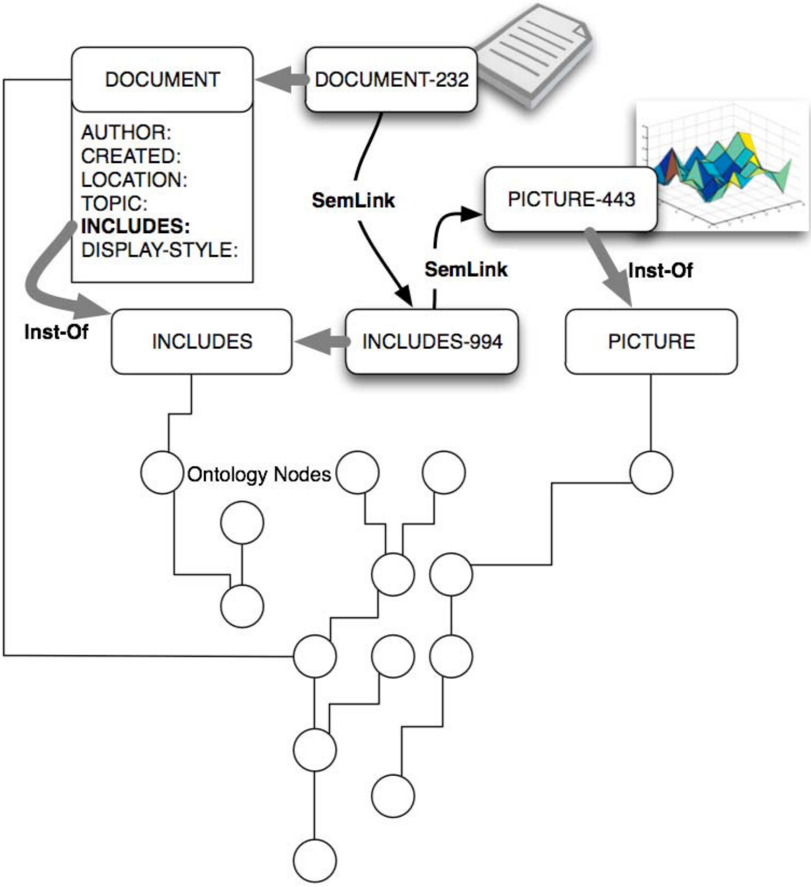


Fig. 1. Metadata for entities, digital or physical, is semantically defined by an underlying ontology that appears to the user as (networks of) SemCards

themselves, leaving only a specification for the desired operations (named processes) to be done on a SemCard in the SemCard itself. This has many benefits, the most important of which is simplicity in usage and ease of maintaining compatibility between systems that use SemCards.

Unique ID. Every SemCard instance has a global, unique identifier (GUID), timestamps representing time of creation and related temporal aspects such as times of modification, as well as a set of policies. Its author is also represented, and any authors of modifications throughout the SemCard's lifetime. The SemCard's policies allow it to be displayed, shared, copied, etc. in predescribed ways, through the use of rules.

Representing any entity. Any type of digital object or information can be pointed to with a SemCard, e.g. a Web page, a product, a service offer, a data record in a database, a file or other media object, media streams, a link to a remote Web service, etc. A SemCard can thus represent any digital item, like a *png* image or *pdf* document, *physical entities* such as a person, building, street, or a kitchen utensil, *as well as immaterial things* like ideas, mythologicical phenomena and intellectual creations. A SemCard can also represent collections, for example a SemCard representing a group of friends would contain links to the SemCards representing the individuals of that social group. Equally importantly, SemCards can represent relationships between SemCards, for example, that a person is the author of an idea.

Display rules. SemCard can carry display rules that dictate how the SemCard itself (as well as its target reference - the thing it represents) should be displayed to the user. These can describe, for example, its owner's preferences or the display device required. As SemCards carry with them their own display specifications their on-screen representation can be customized by their userss; the same SemCard can thus be displayed differently to two different users with different preferences. The rules can specify how metadata and slot values in the SemCard should be organized and what human-readable labels should be used for them, if any, as well as what aspects of the SemCard appear as interactive elements in the interface, and the results of specific interaction with those elements.

3.2 Using Semcards

Creating an instance of a SemCard involves simply selecting the appropriate SemCard template ("template") from a menu or via a search-enhanced selector interface. To fill out the SemCard instance, one or more slots are filled with values – these could be semantic links to other SemCards, typed entities or unclassified content. Each SemCard instance, its semantic dimensions and their values, can be stored as an XML (extensible Markup Language) object, using e.g. the RDF (Resource Description Framework) format [9].

While SemCards could be invisible to users, hidden underneath the standard applications they use, typically a user will want to view and manipulate the information they represent directly, especially for linking them together. For example, a document authored by David, a non-SemCard user, is received by Kris' SemCard system. When received, a SemCard of type "SemDocument" is *automatically* created. Kris links the SemDocument SemCard to his SemCard representing the document's author, David, using an instance of the SemCard type *Authored-By.* This instantly puts the received document in a rich semantic context of the network of all SemCards that link, in one way or another, the document-author pair to a lot of metadata as to who created what at what time, and who shared it with whom, how, and so on.

For viewing and manipulating SemCards we have developed both client-based editors in the spirit of Haystack [10] and Web-based interfaces. Our PersonalRadar desktop application, of which there were made several prototypes, made SemCards actively usable on personal computers, expanding the reach of Twine.com down to personal data, via semantically rich networks. SemCards can be created in many ways; doing so manually from scratch involves selecting a SemCard template type, making an instance of it and customizing its slots using typed entities from an underlying ontology.

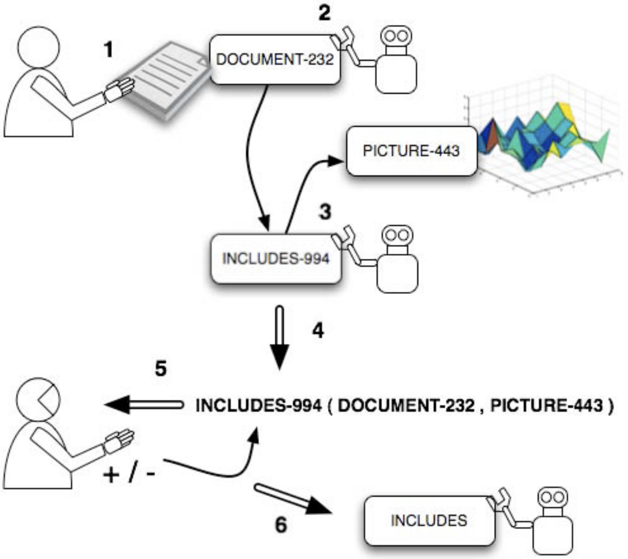


Fig. 2. The iterative nature of human-machine metadata annotation. (1) User creates digital document, (2) a SemCard instance is automatically created; the automation infers that a particular image is included in the document and (3) creates a SemCard for it and a SemCard of type *Includes* that links the two; (4) relationship between the SemCards now forms a triplet that the user can inspect (here shown in prepositional form, but is typically graphical); (5) user modifies the results (+/-) from which (6) the automation processes generalizes to improve own performance.

SemCards templates are ideally fully defined by one or more ontologies. However, the case could arise where a user wants to represent an entity for which no template exists. A user can create free-form slots and collect them into a new SemCard (that has no template). As long as the type of the SemCard – or at least one slot in it – has a connection to a known ontology (it will always have its author and date of creation), the automation mechanisms can use this information to base further automatic refinement of the SemCard instance, like linking it to (what are believed to be) related SemCards. Managing such automatic semantic links becomes akin to unstructured database managment; it will of course never be as good as that for fully-specified SemCards, but because these SemCards live in a rich network of other SemCards, this problem is typically not as large as it may seem.

3.3 End-Users Versus Ontology Experts

In our system expert designers create basic SemCard templates for all major entities such as digital documents, presentations, video files etc., where a template's meta-tags are hand-picked and surface presentation defined (see 3). Importantly, non-experts can then create derivative SemCards by modifying these, adding or removing pre-assigned slots in the SemCards, or making new ones from scratch, using either completely new ones or pre-existing ontologically defined slots (e.g. by copying slots from other SemCards or from a library). All underlying ontological relationships are maintained in the new SemCard; a modified SemCard will store the specifics of its creation history (and can either carry that data around as metadata or link to it in an online database via a GUID). This history information, and its subsequent use and further modification of hundreds or thousands of users, can be used by the automation system to infer about the semantics of the new SemCard and its relation to the underlying ontology, which was not modified in the process.

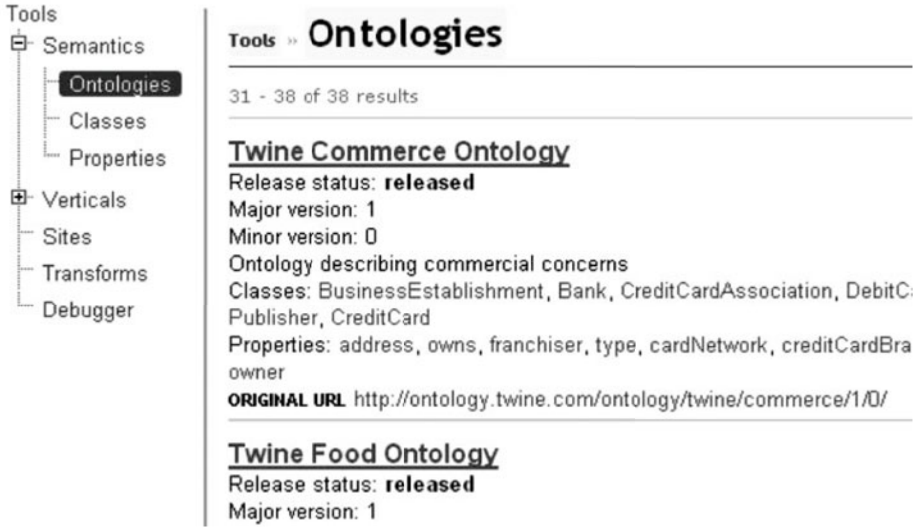
As SemCards isolate the user from the related ontologies, classificatory mistakes in their creation does not destroy the underlying ontologies. This results in a kind of graceful degradation; instead of breaking the system such mistakes only make the automated handling of information in the system slightly less accurate. The relationship between SemCards and the unerlying ontology can be likened to non-destructive editing for video: As the creation history (original data, i.e. ontologies) are not changed but rather represented in a separate intermediate layer, the edit history of any SemCard can be traced back and reverted, if need be, with no change to the underlying ontologies.

Behind each SemCard is thus an ontology that defines the meaning of the SemCard slots, specifies valid values and relations between slots (see Figure 1). An ontology like FOAF (c.f. [11]) or the Dublin Core [12] can be used with SemCards, as each SemCard carries with it a reference to the ontology it is based on. Thus, networks of ontologies can be used with SemCards, whether they use a basic, simple and singleton ontology like the Dublin Core or are definded more deeply in e.g. foundational ontologies such as DOLCE, SUMO [13] [14], or OCHRE [15].

In our current implementation we have created a fairly extensive ontology for important digital data types including Web page, 2-D image, URL, text document, as well as for physical entities such as person, place, organization, etc. The idea is to make this ontology open-source to encourage linking of other ontologies to it, extending its reach and improving its utility, and ultimately bringing the Semantic Web to maturity sooner.

4 Collaborative SemCard Creation by Man and Machine

Through the iterative addition and editing of SemCards by users and the automation mechanisms, a positive feedback loop of iterative improvement on the network is created through such collaboration (Figure 2); initial example networks provide a model for the automation. Reasoning mechanisms are used to infer the implications of corrections to automatically-created data, based on original manual creation. When the initial manual data entry and corrections reaches a critical point the automation starts to provide significant and noticable enhancements to the user. Increased manual input, especially in the form of additions to automatically generated semantic links, allows the automation system also to make inferences about the quality of the data entry, not just for a single user but for many. This allows it to improve the accuracy of its own automation



Tools » Ontologies

31 - 38 of 38 results

Twine Commerce Ontology
Release status: **released**
Major version: 1
Minor version: 0
Ontology describing commercial concerns
Classes: BusinessEstablishment, Bank, CreditCardAssociation, DebitC
Publisher, CreditCard
Properties: address, owns, franchiser, type, cardNetwork, creditCardBra
owner
ORIGINAL URL http://ontology.twine.com/ontology/twine/commerce/1/0/

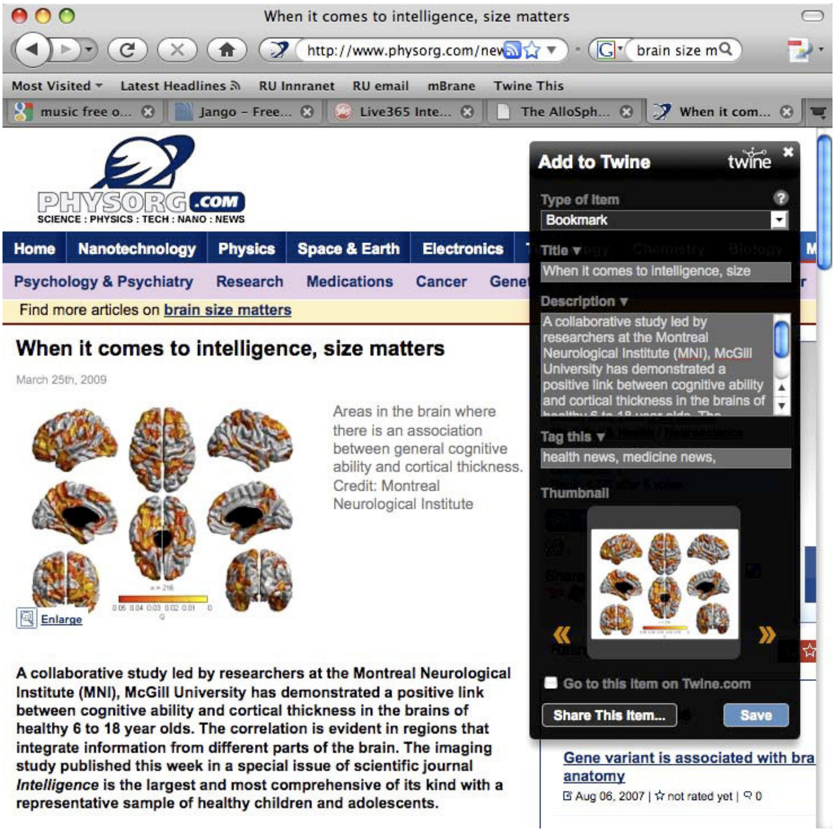
Twine Food Ontology
Release status: **released**
Major version: 1

Fig. 3. The ontology editor allows expert ontology creators to quickly create, manage, connect and extend the multitude of ontologies underlying Twine.com

even further, and suggestions to users about related data will be more relevant and targeted.

An important feature of SemCards is that they record significant amounts of metadata about themselves, including their own genesis. This makes automatic creation of SemCards much more flexible as the automation process can make inferences about the quality of the SemCards (based on e.g. edit history). Because the same representational framework - SemCards - can be used for *all* data, including friend networks, author-entity relationships, object-owner, etc., inferencing can use the multiple SemCard relationship types (e.g. not only who created it but also who the creator's friends are) to decide how to perform automatic relationship creation, data-slot filling, automatic correction or deletion. Moreover, as the SemCard stores its edit history, including who/what made the edits, any such changes can be undone with relative ease. Since this history is stored as semantic information, it can be used to sort the SemCards according to their history. This makes managing SemCards over time much more flexible than if they were history-scarce, like e.g the losely-defined metadata of most data on people's hard drives. For example, caching, compressing or any other processes can be made history-sensitive to a high level of detail.

As an example of collaborative automatic/ manual creation of SemCards, Nova, a SemCard end-user, finds a useful URL and creates a SemCard for it of type "bookmark for a Web-page" (see Figure 4). He makes personal comments on the Web page's contents by making a "Note" SemCard and linking it to the Webpage SemCard. Nova's automation processes, running on the SemCard hosting site, add two things: They fill the Webpage SemCard with machine-readable metadata from the Web page, and they also link these SemCards to a new



Find more articles on brain size matters

When it comes to intelligence, size matters

March 25th, 2009

Areas in the brain where there is an association between general cognitive ability and cortical thickness. Credit: Montreal Neurological Institute

A collaborative study led by researchers at the Montreal Neurological Institute (MNI), McGill University has demonstrated a positive link between cognitive ability and cortical thickness in the brains of healthy 6 to 18 year olds. The correlation is evident in regions that integrate information from different parts of the brain. The imaging study published this week in a special issue of scientific journal *Intelligence* is the largest and most comprehensive of its kind with a representative sample of healthy children and adolescents.

Type of Item: Bookmark

Title: When it comes to intelligence, size

Description: A collaborative study led by researchers at the Montreal Neurological Institute (MNI), McGill University has demonstrated a positive link between cognitive ability and cortical thickness in the brains of

Tag this: health news, medicine news,

Thumbnail

☐ Go to this Item on Twine.com

Share This Item... Save

Gene variant is associated with bra anatomy

Aug 06, 2007 | ☆ not rated yet | 0

Fig. 4. The Twine bookmarklet popup enables Web surfers to create SemCards quickly. The system automatically fills in relevant information ("Title", "Description", "icon", etc.).

SemCard that *it* created, containing further information mined from the Web site. Now Nova shares (a copy) of the SemCard with Jim (it gets saved in his SemCard space), who may add his own comments and links to related SemCards; the fact that the SemCard was shared with Jim by Nova is automatically recorded as part of the SemCard's metadata. Thus, events, data and metadata are created seamlessly and unobtrusively through the collaborative paradigm.

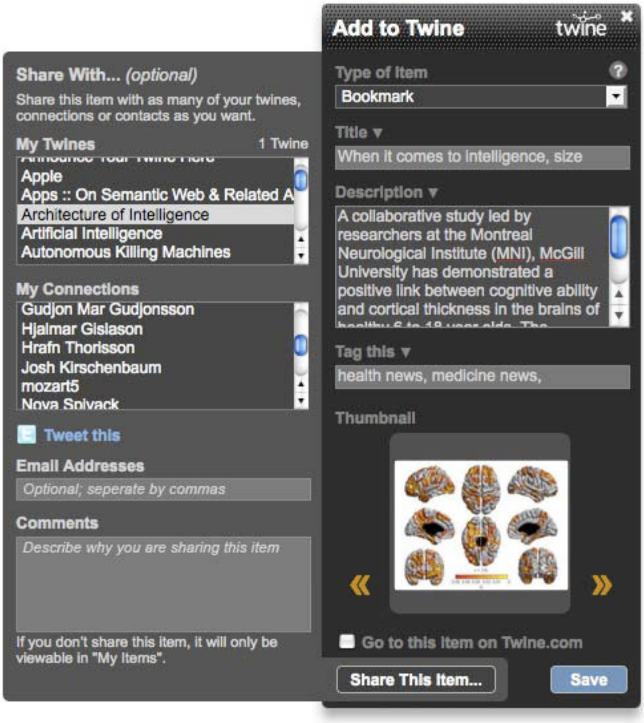
As their authorship is automatically recorded in the SemCards, this can be later used to e.g. exclude all SemCards created by particular automation processes, should this be desired. Proactive automatic mining of a user's SemCards can reveal implicit relationships that the system can automatically make explicit, facilitating faster future retrieval through particular relationship chains in the resulting relationship graphs.

As a users's SemCard database grows user-customized automation becomes more relevant; in the long run, as the benefits of automation become increasingly obvious to each user, people will see the benefits of providing a bit of extra meta-information when they create e.g. a word processing document or an image. This will trigger a positive upward spiral where increased use of automation will motivate users to add more pieces of metadata, which will in turn enable better automation.

5 Deployment on Twine.com

We have implement the SemCard technology and deployed it on the *Twine.com*, an online Semantic Web portal where people can create accounts and use a SemCard-enabled system to manage their online activities and information, including bookmarks, digital files, sharing policies, and more.

As of summer 2009 there were well over 4 million SemCards on Twine.com. At that time Twine.com had around 250 thousand registered users and over 2 million unique visits montly,[3] using the interface developed for the Website. The rate of new SemCard creation had grown to 3K per day, created by an estimated 10% of the users. So far, users seem to rarely correct the automatically-generated SemCards, but a relatively small subset of Twine power users add extensive additional information to them.



Apple
Apps :: On Semantic Web & Related A
Architecture of Intelligence
Artificial Intelligence
Autonomous Killing Machines

Gudjon Mar Gudjonsson
Hjalmar Gislason
Hrafn Thorisson
Josh Kirschenbaum
mozart5
Nova Spivack

Optional; seperate by commas

Describe why you are sharing this item

If you don't share this item, it will only be viewable in "My Items".

Bookmark

When it comes to intelligence, size

A collaborative study led by researchers at the Montreal Neurological Institute (MNI), McGill University has demonstrated a positive link between cognitive ability and cortical thickness in the brains of

health news, medicine news,

□ Go to this item on Twine.com

Fig. 5. Upon creation of the bookmark, SemCard users can choose to share it (left side of popup) with users via Twines they have created or subscribed to ("My Twines"), or directly with people they have connected with ("My Connections")

We will now detail an actual example of making a SemCard for a Web page, a short article on the Physorg.com Web site.[4] As can be seen in Figure 4, when a user comes to a Web page of interest they can click on the bookmarklet "Twine

[3] According to compete.com

[4] `http://www.physorg.com/news157210821.html`

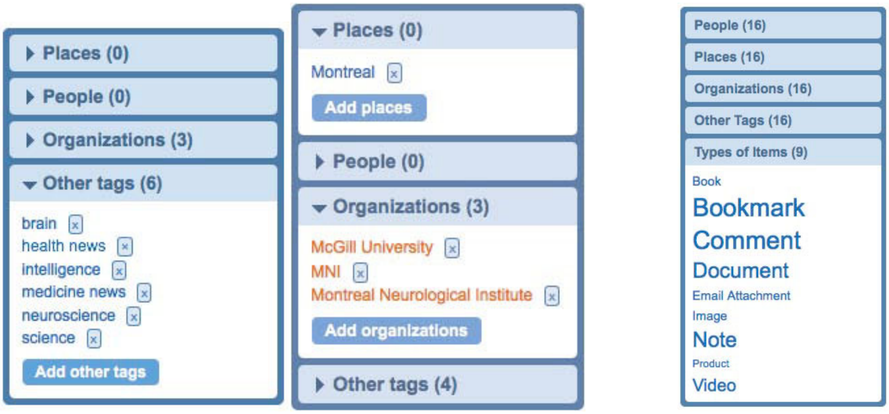


Fig. 6. *Left and center:* Two snapshots of the same dropdown box are shown; automatic processing of a bookmarked Web page can detect places, people organizations and various named entities ("tags"). The user can then modify these by deleting (clicking on the [x]) and adding new ones. The left image shows its "other tags" that were auto-recognized, the middle image showing "organizations", as well as one user-added "place" ("Montreal"). *Right:* Automatic pop-up of items in various categories such as "people", "places", etc., related to a SemCard.

This", which brings up a simple menu with a few information fields. Parts of the SemCard slots have been filled in; the user can choose to edit these, overwrite them with her own information or to leave them as-is. When the user clicks "save" a SemCard for this Web page is created in their Twine account. The user can choose to share this item with users and/or *twines* (see Figure 5) – a twine is a SemCard that can be described as "a blog with controlled access permissions" – in other words, a SemCard for a set of SemCards with particular visual presentation and adjustable viewing permissions.[5] The twine SemCard shows the dynamic properties of SemCards for specifying dynamic processes, e.g. calling on services from mining, inferencing, etc.

When the "bookmark" SemCard is saved, using the "Save" button on the lower right on the bookmarklet popup, the SemCard is stored on Twine.com. Any sharing selection that the user had made during the creation will make the bookmark SemCard available to the users who have permissions to read those twines; for example, sharing it with the twine *Architecture of Intelligence* (Figure 5) will enable everyone who has been invited to subscribe to this twine to see it. In their home page on Twine.com this SemCard will now additionally bring forth a lot of information, including auto-tagging (recognized entities, relationships, etc.).

As seen in Figure 6 a cross next to auto-generated tags allow the user to delete the ones that they don't agree with. Further related information is automatically pulled forward, sorted into "places", "people", "organizations", "other tags" and "types of items": The last one is interesting as it is a unique feature of semantic

[5] We will use "twine" with a lower-case "t" to refer to a SemCard of type "twine".

Webs – here one can find related SemCards of type “video”, for example, or “product”.

Many machine learning techniques can be employed for automatic tagging, entity extraction and relationship detection – in our implementation we have used vector space representations to profile users and their semantic networks and subsequently select related items from other semantic nets. Using a (semantic) drill-down search mechanism a user can further keep refining a search for a SemCard, by selecting any combination of type, tags, author, etc. (Figure 7). During such drill-downs, suggestions by the automation of related material become increasingly better.

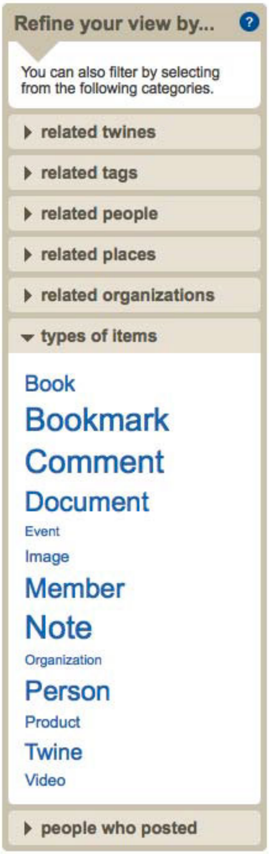
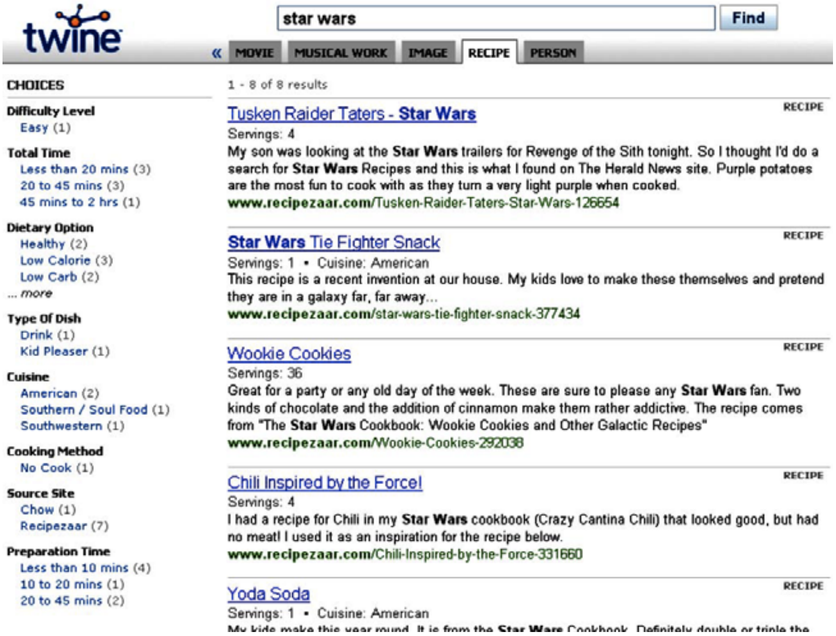


Fig. 7. One type of semantic search box on *Twine.com*

6 Future SemCard-Driven Automation Services

As already mentioned, the system we have developed enables automatic semantic mining of content sent by a user. Using existing semantic networks created manually and automatically by the system, this mining can be done without requiring any actions or special editing by the user, such as inserting special characters or identifying terms or phrases as potential semantic objects. We will now provide a few examples of future automation services enabled by the SemCard technology. These have not been implemented as services yet, but prototypes already exist.

Intelligent E-Mail/Sharing. An example of a potential future use of the SemCard technology is for email-like purposes. In this example the user has a semantic email account with a semantic service provider (or the user keeps the same normal non-semantic email account but adjusts mailbox settings so that mail received and sent are processed by the provider). The semantic service provider processes all incoming and outgoing e-mail, automatically creating SemCards representing the e-mail itself and concepts referenced in the email and identified by entity detection algorithms [16]. No intervention would be required of the user, other than initial set-up. When the SemCards have been created they are automatically linked to other previously defined semcards in the user’s account, enlarging the user’s knowledge network. Now the user can for example find all emails “sent by Jim to Nova about Semantic Web technologies regarding the PersonalRadar product” – a perfectly valid search using semantic relations built from information readily available in the user’s account.



star wars

1 - 8 of 8 results

Tusken Raider Taters - **Star Wars**
Servings: 4
My son was looking at the **Star Wars** trailers for Revenge of the Sith tonight. So I thought I'd do a search for **Star Wars** Recipes and this is what I found on The Herald News site. Purple potatoes are the most fun to cook with as they turn a very light purple when cooked.
www.recipezaar.com/Tusken-Raider-Taters-Star-Wars-126654

Star Wars Tie Fighter Snack
Servings: 1 • Cuisine: American
This recipe is a recent invention at our house. My kids love to make these themselves and pretend they are in a galaxy far, far away...
www.recipezaar.com/star-wars-tie-fighter-snack-377434

Wookie Cookies
Servings: 36
Great for a party or any old day of the week. These are sure to please any **Star Wars** fan. Two kinds of chocolate and the addition of cinnamon make them rather addictive. The recipe comes from "The **Star Wars** Cookbook: Wookie Cookies and Other Galactic Recipes"
www.recipezaar.com/Wookie-Cookies-292038

Chili Inspired by the Force!
Servings: 4
I had a recipe for Chili in my **Star Wars** cookbook (Crazy Cantina Chili) that looked good, but had no meat! I used it as an inspiration for the recipe below.
www.recipezaar.com/Chili-Inspired-by-the-Force-331660

Yoda Soda
Servings: 1 • Cuisine: American
My kids make this year round. It is from the **Star Wars** Cookbook. Definitely double or triple the

Fig. 8. The user interface for searching large collections of SemCards has a familiar, easily navigated multi-column tabbed layout

Because the underlying representation for this sharing of text messages is the SemCard, this activity constitutes *semantic sharing*. Its principles can be applied to *any* digital object – with a SemCard client the same sharing method used for the email SemCards can be used for sharing any digital object; there is no need to use "attachments" for sharing such entities as they are first-class objects with full meta-data about their history including creation, manipulation and sharing events.

Semantic social networks from emails. Another important feature enabled by SemCards is the creation of semantic relationship networks, where a user's relationships is automatically mapped based on email correspondence. Such a network will be most useful if the correspondence is also based on SemCard technology, as described above, but regular email can also be used to form the basis for this technology. To move to this technology from their current software, users provide their private and business contacts in their account, for example by uploading their address book into the system.

Depending on various factors, including the content and number of the emails exchanged, these relationship link types include types such as "friend", "colleague", "relative", "conversants", with the last type being the most generic. To take the example of the "conversants" link type: The link is created between the user and another person when they have exchanged at least two emails, where the second email was a response to the first. The link contains the time of the exchange (time of sending, time of reading, both emails), as well as who made the link, and when (even when automatically created). As before, the user can set a preference for minimal, medium, or heavy mining of her email.

The system processes the addressees of all emails to infer who is communicating with the user about what, as in the above email example, as well as inferring with whom the user has relationships, what kind of relationships those are, and what projects they relate to. Emails are then linked to those inferred projects. This enables very powerful personalization of information displaying: For example, sets of different preference settings can be associated with separate (named or unnamed) groups of contacts, enabling differential treatment depending on who the user communicates with. A group called "friends" may have certain settings for how entries from/to them should be formatted for viewing; a "personal Facebook-like service" with a corresponding look could be set up by a user for one of her groups, while using a vastly different display setup for others.

7 Conclusions

To realize the full potential of the Semantic Web vision, several challenges must be met. One of these is the unreliability of automated metadata creation systems, another is the lack of a strong and flexible framework for representing data and metadata. We have developed SemCards, which solve these challenges in a way that takes advantage of current technologies while allowing for future growth in the foreseeable future. We have implemented this technology on the desktop as well as on the Web, showing it to scale to hundreds of thousands of users.

The technology presents a powerful representation scheme that enable collaborative human-machine and human-human creation of Semantic Web information. SemCards achieve this by separating hard-core ontologies from the end-user, mediating these via graphical information structures, represented under the hood using RDF and OWL, while supplying their own visual representation schemas for on-screen viewing. The SemCard framework allows better sharing, storing, annotating, enhancing and expanding semantic networks, creating true knowledge networks through a collaboration between people and artificial intelligence programs.

The Semantic Web site *Twine.com*, which has well over 2 million monthly unique visitors, has demonstrated the usefulness and extensibility of the technology. In close collaboration with automation processes, these users have created over 5 million SemCards to date. Our results so far show that SemCards can support all of the features described in this paper for over 300 thousand users and we have good reason to believe that the technology will scale well beyond this.

Other proposed approaches for realizing the Semantic Web vision have fallen short on one or more of the key features that SemCards address and solve. We believe that as a uniform standard for representing data and metadata on the World Wide Web, SemCards, or a related technology, could very well be the missing glue that is needed to link together the forces – natural and artificial – that are needed to propel the Web forward to the next level, the semantic level.

Acknowledgments. The work was supported by Radar Networks, Inc., by Reykjavik University, Iceland, and by the School of Computer Science at Reykjavik University. The authors would like to thank the Twine team for the implementation of the SemCard technology on Twine.com, as well as the funders of Radar, Vulcan Ventures Inc., Leapfrog Ventures Inc. and the angel investors.

References

1. Dill, S., Eiron, N., Gibson, D., Gruhl, D., Guha, R., Jhingran, A., Kanungo, T., Rajagopalan, S., Tomkins, A., Tomlin, J.A., Zien, J.Y.: SemTag and Seeker: Bootstrapping the Semantic seb via automated semantic annotation. In: Proceedings of the World Wide Web Conference (2003) doi: 10.1145/775152.775178
2. Etzioni, O., Gribble, S.: An evolutionary approach to the Semantic Web. Poster presentation at the First International Semantic Web Conference (2002)
3. Maly, K., Nelson, M.L., Zubair, M.: Smart objects, dumb archives: A user-centric, layered digital library framework. D-Lib Magazine, 5 (1999)
4. Kahn, R., Wilensky, R.: A framework for distributed digital object services. International Journal on Digital Libraries 6, 115–123 (1995)
5. Payette, S., Lagoze, C.: Flexible and extensible digital object and repository architecture (FEDORA). In: Nikolaou, C., Stephanidis, C. (eds.) ECDL 1998. LNCS, vol. 1513, pp. 41–59. Springer, Heidelberg (1998)
6. Lagoze, C., Payette, S., Shin, E., Wilper, C.: Fedora: An architecture for complex objects and their relationships. Journal of Digital Libraries - Special Issue on Complex Objects (2005) doi: arXiv:cs/0501012v6
7. Karger, D.R., Bakshi, K., Huynh, D., Quan, D., Sinha, V.: Haystack: A general-purpose information management tool for end users based on semistructured data. In: CIDR, pp. 13–26 (2005)
8. Drummond, N., Jupp, S., Moulton, G., Stevens, R.: A practical guide to building OWL ontologies using the Protégé 4 and CO-ODE tools, 1.2. edn. (2009)
9. Decker, S., Melnik, S., Harmelen, F.V., Fensel, D., Klein, M., Erdmann, M., Horrocks, I.: Knowledge networking in the Semantic Web: The roles of XML and RDF (2000)
10. Huynh, D., Karger, D.R., Quan, D.: Haystack: A platform for creating, organizing and visualizing information using RDF. In: Semantic Web Workshop, WWW 2002 (May 2002)
11. Ding, L., Zhou, L., Finin, T., Joshi, A.: How the Semantic Web is being used: An analysis of FOAF documents. In: Proceedings of the 38th International Conference on System Sciences, p. 313.3 (2005) doi: 10.1109/HICSS.2005.299
12. Sutton, S.A., Mason, J.: The Dublin Core and metadata for educational resources. In: International Conference on Dublin Core and Metadata Applications, pp. 25–31 (2001)
13. Varma, V.: Building large scale ontology networks. In: Language Engineering Conference (LEC 2002), Hyderabad, India, p. 121 (2002)
14. Hong, J.F., Li, X.B., Huang, C.R.: Ontology-based predication of compound relations: A study based on SUMO. In: Proceedings of PACLIC18, Waseda University, Japan, pp. 151–160 (2004)
15. Schneider, L.: Designing foundational ontologies: The object-centered highlevel reference ontology OCHRE as a case study. In: Song, I.-Y., Liddle, S.W., Ling, T.-W., Scheuermann, P. (eds.) ER 2003. LNCS, vol. 2813, pp. 91–104. Springer, Heidelberg (2003)
16. von Brzeski, V., Irmak, U., Kraft, R.: Leveraging context in user-centric entity detection systems. In: Proceedings of the 16th Conference on Information and Knowledge Management (2007)