From folksologies to ontologies: how the twain meet

Peter Spyns, Aldo de Moor, Jan Vandenbussche, and Robert Meersman

Vrije Universiteit Brussel - STAR Lab,
Pleinlaan 2 Gebouw G-10, B-1050 Brussel - Belgium
(Peter.Spyns|Aldo.De.Moor|javdbuss|Robert.Meersman)@vub.ac.be

Abstract. Ontologies are instruments for capturing and using formal semantics, and are often the result of a "central committee controlled" style of working. A new trend on the Web is the increasing popularity of folksologies in the form of social bookmarking sites. Folksologies provide informal semantics and can be created and adopted by anybody anytime anywhere on the Internet. Shared meaning in a folksology emerges through the use of tags that are used to bookmark web pages, their usage frequency being considered a reliable indicator of their usefulness and acceptance.
Rather than choosing for either ontologies or folksologies, hybrid emergent semantics systems are needed that combine elements of both perspectives, depending on the particular application. There is a need to analyse the larger picture (including the full range of semantics' functionalities in their context of use.
In this paper, we outline a number of key design characteristics of emergent semantics systems (ESS). We examine the functionalities of two existing examples of well-known ESSs: del.icio.us and Piggy Bank. Using the results of this comparison, we introduce DogmaBank as a proof of concept implementation of a next-generation ESS that introduces a more advanced combination of lexical and conceptual emergent semantics functionalities.

1 Introduction

Many definitions of ontologies exist [18, 20, 58, 57]. We prefer the one from Guarino [22, p.7][1]. Combining various definitions, an ontology can be seen as a formal, shared, explicit but partial specification of the commonly agreed upon intended meaning of a conceptualisation. With some form of simplification, one could say that an ontology is like a dictionary with unambiguous meaning defining entries linked by many formal relationships. A taxonomy on the other hand has only a single structuring relationship between its constituents - e.g., is_a_kind_of, is_a_part_of.

The term 'folksology' has been coined by Stefano Mazzocchi [36] in analogy with 'folksonomy', created by Thomas Vander Wal [2]. It basically means that ontologies resp.

[1] An ontology is a logical theory accounting for the intended meaning of a formal vocabulary, i.e. its ontological commitment to a particular conceptualisation of the world. The intented models of a logical language using such a vocabulary are constrained by its ontological commitment. An ontology indirectly reflects this commitment (and the underlying conceptualisation) by approximating these intended models

[2] http://en.wikipedia.org/wiki/Folksonomy

taxonomies emerge on the fly thanks to individuals who autonomously create and use any tag they like to annotate any web page they deem worthwhile [3]. The tag as well as the tagged web page (URL) is published in a repository (= a social bookmarking site). Anybody can also re-use any tag from anybody else if it has been published on the social bookmarking site. Within what is called the Web2.0 community these social bookmarking sites are becoming very popular [25]. Flickr [4], an image sharing site, and del.icio.us[5], a bookmark collection site are seen as the two most successful representatives of such sites. In the academic world, there is, for instance, Bibsonomy [6], a site to share bibliographic references.

As a consequence, some (e.g., Shirky [50]) even hail folksologies as the best way to create meaning on the Internet as opposed to ontologies as defined in the tradition of Gruber [18] and Guarino [20]. Note that an ontology by definition is a social construct as it constitutes a shared agreement resulting from some form of meaning negotiation by the various domain stakeholders, although this social aspect in practice is often at best assumed by ontology engineers - in many cases acting on their own.

For those familiar with the world of database schema modelling, the current social bookmarking services closely correspond, albeit it in a more global and networked environment, to the situation in which a database schema modeller invents his own table and column names and expects other modellers to understand them as such. The shared meaning, however, only resides in the brain of each individual that uses and tries to understand such a column name (or tag in the folksology case), as it basically equals a natural language word. As soon as multiple database applications are supposed to interoperate with each other (federated databases, data warehouse, ...), problems arise as more often than not the meaning of the labels has not been made explicit. Well-known issues like ambiguity, cryptic abbreviations, multilinguality etc. inhibit database interoperability and are bound to also manifest themselves in the area of folksologies as the many people involved in creating and using the tags have different mental models.

Much research effort in the database community has been dedicated to schema matching, schema merging, DB-mediators (e.g., [3, 28]) and so on to solve or at least reduce interoperability problems. Ontologies are recognised by many as instrumental for solving the interoperability problem as they are basically a meaning repository for labels and, in principle, independent of any specific application [38].

1.1 Bridging the gap

Folksologies and ontologies are not two opposite ways of organising a (more) meaningful Internet, but rather constitute two ends of a range. The strenghts of one approach can

[3] There is a lot of confusion about and inconsistent use of both terms. Folksonomies relate to folksologies in the same way as taxonomies relate to ontologies, i.e. a single vs. multiple types of relationships between the concepts. Their commonality, however, is the focus on the social process leading to informal semantics. In the remainder of this paper we will use the term 'folksology', since that is the clearest counterpart of ontology.

[4] http://flickr.com/

[5] http://del.icio.us

[6] www.bibsonomy.org

offset the weaknesses of the other. Shirky distinguishes some characteristics that makes a domain more (or less) suited for an ontology to be built for it [50] [7]. In the same way, he contrasts the participants of ontologies and folksologies. Table 1 lists these two aspects.

characteristics	*ontology*	*folksology*
corpus	small	large
categories	formal	non formal
entities	stable and restricted	non stable and unrestricted
edges	clear	unclear
paticipants	*ontology*	*folksology*
users	coordinated and expert	uncoordinated and amateur
authority	authoritative sources of judgment	no authority
cataloguers	expert	naive

Table 1. ontologies vs. folksologies: their characteristics and participants [50]

Whereas Shirky overall favours folksologies for real-world applications, Gruber makes the case that ontologies have an important role to play as well in supporting the tagging process. This view is reflected in the TagOntology project, which is about identifying and formalizing a conceptualisation of the activity of tagging, and building technology that commits to the ontology at the semantic level [19].

Although the debate is ongoing about the pros and cons of folksologies versus ontologies, still little attention is paid to the next step: what do the *systems* look like that are going to be built on top of these approaches? We define *emergent semantics systems* as information systems that combine informal and formal semantics approaches (i.e. folksologies and ontologies) to optimally serve the evolving requirements of communities of human and machine information providers and users. To support our points, we have implemented a proof of concept ESS combining a social bookmarking tool with ontology technology. It should help to build bridges not only between these informal and formal semantics approaches, but also between the theoretical research and practical systems development communities. In this respect, this paper can be seen as a position paper complemented by a proof of concept implementation.

1.2 Structure of the paper

The remainder of this paper is organised as follows. The next section treats the notion of emerging semantics theory and systems in more depth. In section 3, we outline six

[7] Shirky sees an ontology mainly as a (library) classification system with, for instance, the Yahoo directory as its Internet version, which leads Gruber to state that "Shirky uses 'ontology' where he should have used 'taxonomy', and therefore he "misses the point [on ontologies] ... so beautifully [19]. See also [41] for a rebuttal of Shirky's opinions. Note that in general statements in the folksologies vs. ontologies debate represent individual positions that have not been submitted to scientific peer review.

characteristics that need to be taken into account in the design of an ESS. In section 4, we illustrate our ideas on two existing Web2.0 systems (del.icio.us - section 4.1 - and Piggy Bank - section 4.2) as well as propose a new system for semantic annotation (section 4.4) based on the DOGMA ontology framework (section 4.3). We end the paper with a section on related and future work - section 5 - and a conclusion (section 6).

2 Emerging Semantics: From Theory to Systems

Research on the Semantic Web and ontologies posits that these technologies are meant to be used by *software agents* (rather than human agents) that offer and look for all kinds of services [5]. For these kinds of Semantic Web applications, a formal and explicit definition of meaning shared by these agents is prerequisite as they are expected to reason about or infer new knowledge. For a domain of discourse or application domain a conceptualisation is created and later on implemented as an ontology [22]. Consequently, many resources are spent focusing on defining (semantic) web (service) languages and (reasoning) formalisms. This in contrast to social bookmarking services that are more tailored towards humans interacting. As an implicit consequence, the Semantic Web seems to be primarily oriented towards business rather than social use. Many of the classical Semantic Web motivating examples involve humans, but through the use of software agents, merely as individual consumers or clients who book travels, buy books and CDs, use smart house-hold devices and enjoy health monitoring and personal scheduling services. Typical usage of current Web2.0 applications, in particular community building and social aspects, is less considered by the Semantic Web research community. On the other hand, the Semantic Web interoperability topics, crucial for the intelligent software agents that will ease our on-line life, are hardly valued (or even plainly critised e.g., [49]) by the Web2.0 community.

A novel topic in the Semantic Web domain is called "Emergent Semantics". Although the original theme addressed by Emergent Semantics was rather restricted to database query meaning transformation and preservation in a peer-to-peer setting [2], the meaning of the expression has been widened to the result of dynamic and distributed local processes during which new or additional semantics are obtained and applied on a global level [8]. Another application domain concerns multimedia annotation (e.g., [17]). Aberer et al. [1] discuss the following characteristics of emergent semantics. It concerns agreements that

1. are a semantic handshake protocol
2. emerge from negotiations
3. emerge from local interaction
4. are dynamic and self-referential approximations
5. induce semantic self-organisation

In that respect, folksonomy and folksology techniques and software facilitate the emergence of distributed semantics. The semantics emerge from the implicit but immediate feedback from the community in the form of usage frequencies of tags as well as

[8] Conclusions of the scientific meeting of the IFIP WG 2.6 on Databases at 30/10/2005 - see [8].

the listing of tags per URL and URLs per tag. A tag creator can conform his usage of tags to the "average", and thus implicitly agrees on the meaning attributed to a tag.

Our overall position is that there will many emergent semantics systems consisting of multiple semantic webs and light-weight ontologies that are "valid" for a specific application or business domain, (professional) organisation and the like. These semantic webs will include social aspects in their creation, maintenance, and evaluation stages. Nowadays company-wide controlled vocabularies (e.g. within Boeing and General Motors), and even world-wide classification systems for a specific domain (e.g., the 10th edition of the International Classification of Diseases) exist, are permanently used and regularly updated. As ontologies are, said in a simplifying way, formal representations of such vocabularies and classifications, there is in principle no reason why the transition to the Semantic Web of these existing mechanisms of establishing meaning agreements should not work. On the other hand, ontology engineers should have the candour to admit that this transition phase has not yet been studied thoroughly.

In a nutshell, folksologies have to differentiate between natural language words and a language-independent artificially created label indicating a concept or sense, and ontologies have to adopt strategies and tools as used for social bookmarking sites to make the meaning of concepts spontaneously emerge and converge.

3 Characteristics of Emergent Semantics Systems

In this section, we describe six key characteristics of emergent semantics systems including human agents. The core issue is how informal and formal meaning emerges, is agreed upon and shared by a community, and is used subsequently for various, including unforeseen, scenarios on a global or potentially world-wide scale by human and software agents that can organise themselves in (virtual) communities or societies. The choice of characteristics has been inspired by contrasting essential features proposed by proponents of the Semantic Web and the Web2.0, respectively. For each characteristic, we sketch how the informal and formal semantics approaches do or could meet.

3.1 Natural language words vs. language independent concept labels

Folksologies are based on a very basic but implicit assumption: everybody speaks the same language - in practice US English. The meaning agreement happens unconsciously: individuals understand what a tag means since the tag is a single[9] plain US English word [10]. We've already mentioned synonyms and homonyms, but what with other languages ? E.g., is a Dutch speaking individual simply expected to use US English to conceptualise his small world ? What happens if somebody uses a wrong or inappropriate translation ? Maybe somebody doesn't understand the meaning of a tag ? A word lexicalises a concept that is an idea, notion or meaning (cf. the meaning triangle of Ogden and Richards [45]). There is an m:n mapping of words on concepts, i.e. one

[9] Del.icio.us does not allow compound words - e.g., 'semantic web' is considered as two tags.

[10] We avoid to use 'term' as it means logical term on the one hand or technical term of a special purpose (natural) language on the other.

word expresses several meanings (homonym) or several words express the same meaning (synonym)[11]. In addition, according to the context and language, a concept can be expressed differently (e.g., fish is translated in Spanish by 'pez' or 'pescado' depending on whether or not the fish is still living ('pez') or on your plate to be eaten ('pescado') [12]. The point is that one should not use words to label concepts but rather unique identifiers that stand for language independent (or neutral [43]) concepts to which words are mapped [54]. Of course, this requires that the meaning uniquely identified by a tag is explicitly defined - cf. e.g., WordNet synsets. Meaning definitions can come from dictionaries, encyclopedias, company controlled vocabularies etc. An important consequence is that additional software support is needed to avoid that duplicate and redundant tags are created and to steer a tag user into selecting the most appropriate tag. Unfortunately, also ontology engineers often fail to make this fundamental difference between a word and a concept [43].

3.2 Informal vs. formal semantics

Folksologies suffer from the fact that they stay on the language level and thus are unable to cope with issues such as synonymy, notational variants and homonymy [13]. Ontologies (formal models) on the other hand are unable to "connect" to the real world as they are represented in some first order language [14, pp.30-31] [14].

Informal and formal semantics can, for example, be aligned resulting in explicit meaning circumscriptions that are linked to a concept identifier or label (= the logical term) and that group various lexicalisations (= the language term) depending on the context and language [54]. A concept label is a unique identifier for a synset or sense (as is the case for (Euro)WordNet). Formal semantic restrictions (e.g., expressing uniqueness or mandatoriness) and reasoning rules (e.g., defining transitivity for a certain relationship) use the concept identifiers as logical terms. In most of the cases, a system internal (alpha-)numerical ID exists next to a more human-friendly pseudo-word string. The signature of an ontology is now explicitly rooted in "natural" semantics (e.g., by means of dictionary definitions), while at the same time meaning ambiguity is avoided (a tag is unambiguously linked to a specific sense). Thanks to the additional formal semantic restrictions and definitions, the fuzziness of informal natural language definitions is further reduced. Formal definitions are also needed for inferencing.

3.3 A bag of loose words vs. a well structured semantic network

As already mentioned earlier, a folksonomy consists of words. In addition, there is a complete absence of any structuring relationship. Some (e.g., [34]) consider that the

[11] See [34] for some "life examples" from Del.icio.us regarding ambiguity and synonyms.

[12] Cf. also the difference in U.K. English between 'pig' and 'pork' (but in Dutch there is only the word 'varken').

[13] Simply adding an OWL "SameAs" or "DifferentFrom" statement, as suggested by Mazzocchi [36] is insufficient as it only marks an equivalence/difference between two language terms but does not provide any explicit definition of the meaning of the tag.

[14] Model-theoretic semantics does not pretend, and has no way to determine what certain words and statements "really" mean. (...) It [= model theoretic semantics] offers no help in making the connection between the model (the abstract structure) and the real world.

absence of a hierarchy to agree upon favours cooperation for social bookmarking. In fact, the term 'folksonomies' is semantically speaking too broad as there is no taxonomy involved, but only a vocabulary (a bag of words) [15]. 'Folksabularies' (in analogy with vocabularies) would thus be a more accurate term. Relationships are a fundamental part of ontologies. A classical relationship is the 'subclass_of' relation. Description Logics reasoners rely on these relationships. Ontologies allow for the formal representation and analysis of many other types of relationships and constraints as well, thus allowing, for instance, for advanced retrieval services on large knowledge bases

3.4 Effort at creation time vs. effort at usage time

As usual, there is a trade-off: the higher quality results have to be, the more (preparative) time and resources it takes. A folksology is relatively easy to build (anybody is able to create tags or use anybody else's tags), but at the cost of potential misunderstandings and noise at browsing and retrieval time for which an individual user has to compensate by evaluating and selecting the look-up results himself. An ontology requires a substantial effort of experts on reaching meaning agreements and formally representing them, but then has the advantage of setting the semantics as precise as possible and grounded in the domain, which should remove potential ambiguities and misunderstandings in favour of interoperability. The situation is comparable to the classical word-based search engines that produce a large amount of noise. Semantics-based search engines, such as OntoSeek [21] and Swoogle [12], offer the promise of returning more accurate results and reducing noise. Typically IE/IR results are measured in terms of recall (how many elements of the ideal answer set are contained in the actual answer set?) and precision (how many elements of the actual answer set are correct ?). It is widely assumed that indexing pages with ontology tags will improve IE/IR results. Likewise, bookmarking pages with tags for which there exists an agreement and explicit definition of their meaning will result in more accurate look-up results.

3.5 Individual creation decision vs. group-wise creation agreement

The fact that in a folksological setting, any individual is the allmighty "deus ex machina" who can very easily create, rename, group, split, merge and delete tags and thus categorise and classify his small universe at will largely explains the success of folksonomies - see e.g., [34]. In addition, by making use of social bookmarking sites and their software, the individual receives an immediate feedback on how other individuals use (the same) tags (e.g., all sites bookmarked by the 'tools' tag are listed by browsing http://del.icio.us/tags/tools) and can adapt/conform his tagging behaviour. Although in principle, ontologies should be the result of a group-wise negotiation process, the current practice unfortunately still is that a single ontology engineer encodes his conceptualisation of a domain (or in many cases an application) in an ontology language such as OWL and subsequently hopes that other stakeholders in the domain are happy with it and adopt it. The immediate feedback loop of folksologies is missing here. In order

[15] Del.icio.us allows a user to define a label regrouping tags, but the system then only uses it for local display purposes.

to support the agreement process that links a unique label to a meaning description, a distributed negotiation and decision support system and logging mechanism is needed. The suggestion of using folksologies as the start for professional controlled vocabularies ([40]) should be retained [16] and could serve in the same way as quantitative linguistic corpus analysis does to determine the most important terms for a domain.

3.6 The (unbearable?) lightness of being a tag creator vs. the authoritative weight of a relevant stakeholder

As anybody is allowed to create a social bookmark, no single individual is able to reliably assess the value of the tags. Only the frequency of use gives an indication of how others perceive the relevance and value of a tag. On the Internet, all tag creators are equal. But also here, some are more equal than others. Some persons simply are more knowledgeable and have better expertise so that their view on a part of the world is accepted as more relevant, better organised, etc. Not just anybody should participate in a committee of representative experts and prime stakeholders that decides on the content of an ontology. In a professional environment this way of working is widely applied to build controlled vocabularies or decide on the use of XML tags, for instance. The creators of an ontology are attributed an authoritative weight that inspires trust to potential users of this ontology about its quality, especially as objective measures to evaluate ontologies - see e.g. [55] - are still in their infancy [6, 24]. Web pages annotated with qualitative tags will be preferentially visited by intelligent software agents. Merely looking at frequencies of tag use is too simplistic a measure [17]. One could attribute weights to authoritative users, but this implies a preliminary agreement (and limited in time ?) on the weigths attributed to these authorities. We prefer a combination of authoritative weights and frequency of usage as, for instance, has been implemented in the FolkRank algorithm[26]. In the areas of cultural anthropology and group communication theory, interesting research on consensus analysis is done taking into account the authority of an information source [4].

4 Implementing an Emergent Semantics System

In this section, we describe how a social bookmarking system has been combined with an ontology server in order to create an ESS along the lines as described in the previous section and summarised by the following requirements:

- differentiate words from senses
- combine formal and informal semantics
- structure the domain of discourse
- combine individual suggestions with committee style decisions
- combine spontaneous trends with expert opinions

[16] However, the fact that only simple words are used as tags seriously biases the outcomes.
[17] E.g., one cannot simply state that a definition is incorrect only because it is hardly used.

Before discussing the actual implementation (section 4.4), we summarise the characteristiscs of two existing social bookmarking tools, namely del.icio.us (section 4.1) and Piggy Bank (section 4.2) to mark their inadequateness. Our system called DogmaBank extends the Piggy Bank open source software [18] with DOGMA technology (section 4.3) in order to meet the requirements mentioned above.

4.1 Del.icio.us

Del.icio.us is representative of most social bookmarking sites - other well known ones are Connotea [31] and Bibsonomy. Any human (tagger) is able to create, rename, delete, merge and split tags (= a single word) and associate them with a web page. The system stores a tag, its associated URL and tagger and makes this information available to any other user (subscriber) who can annotate his web pages with these tags. Relations between tags are either based on frequencies ("Related tags") or aggregates of tags (using the "Bundle" functionality to create containers of tags).

Homonyms nor synonyms can be detected at creation time. Only when a subscriber examines with which pages a certain tag is associated he discovers potential cases of homonymy and synonymy. No informal or formal definitions are provided. As there is no possibility to create conceptual representations, only human agents can make use of this tool. There is no explicit group interaction foreseen (only implicitly as via usage frequencies one sees how the group behaves).

4.2 Piggy Bank

Piggy Bank [27] is another, more recently developed, social bookmarking tool that disposes of an extended functionality compared to del.icio.us, especially regarding the sharing, organising and re-using of the information collected during bookmarking (and scraping) activities. A user creates a new tag (= keyword) but "under the hood" a URI to an RDF resource (with the keyword as the OWL label) is created. However, the user interface does not yet support applying the OWL:DifferentFrom and OWL:SameAs primitives to cope with synonymy and homonymy.

Nevertheless, except for the merely syntactic definition of the RDF resource, no explicit formal or informal description of its exact meaning is available. As a consequence, only human agents are intended to make use of this tool. But at least, the Piggy Bank system designers are aware of the difference between a keyword and a concept and have foreseen an initial mechanism for it - although not yet applied in practice. It implies that no relationships between the tags (or rather the keywords pointed to) are available, except for a listing of URLs per tag.

4.3 DOGMA

The DOGMA (**D**eveloping **O**ntology-**G**rounded **M**ethodologies and **A**pplications) ontology engineering framework developed in our group builds on the fundamental distinction between an ontology base and a commitment layer [29]. This double articulation [51] in an intuitive ontology base and formal commitment layer (formal semantic

[18] Piggy Bank is licensed under the BSD license - see http://simile.mit.edu/piggy-bank/ .

constraints on a selection of the represented domain knowledge) is combined with basic insights from linguistics to offer a powerful framework for ontology engineering and modelling [39] that takes into account the difference between a natural language word and a concept label (logical term) identifying some notion or concept [54]. To cope with homonyms and synonyms we state that a concept is lexicalised for a certain language in a certain context by a specific word[56] [19] (cf. the famous example of the many meanings of "bank" [7] or the various denominations for beer glasses in Australia [23, p.206]. As the DOGMA Concept Definition Server implementation incorporates the principles mentioned above, its functionalities (e.g., one method returns the concept label associated with a particular term) constitute a good basis to add a conceptual layer to Piggy Bank. To our knowledge, there are not so many ontology servers that distinguish between a natural language term and a concept - another one is for example KAON-2[20].

Methods and practices of distributed collaboration and negotiation for ontology engineering are an important aspect of ontology engineering - [11] describes a collaborative approach within the DOGMA framework. Not only "committee-style" approaches should be looked at, but also spontaneously emerging meaning agreements. Especially for the latter scenario we hope to incorporate some of the "folksological" techniques next to methods from social sciences, possibly supported by ontologies [10].

4.4 DogmaBank

In the following, we outline how we have (partially) complemented and extended the Piggy Bank software with an ontology server that is able to link words or terms to concepts and vice versa. Remember that the del.icio.us and Piggy Bank lack any conceptual machinery and collapse words and meaning but offer an intuitive way to create annotations, while the DOGMA technology should include means to cater for the phenomenon of emergent meaning definition in its formal conceptual engineering framework.

We could have chosen to build YABT (Yet Another Bookmarking Tool) from scratch, using one of the rapid Web2.0 development frameworks that are increasingly becoming available. However, we feel this would have been an isolated attempt. Instead, we have adapted a well known system, Piggy Bank (see section 4.2), of which we preserved all of the existing functionality in our DogmaBank implementation. This makes the transition easier for a Piggy Bank user. We have implemented a new user interface to support *ontology-based tag creation and tagging* processes (the upper part of Figure 1). This interface replaces the standard Piggy Bank interface. It is implemented as a Firefox plug-in [21]. The plug-in connects to the DOGMA Concept Definition Server to retrieve WordNet [22] (or other WordNet-style) senses that are associated with the keyword entered (the right part of Figure 1). We refer for the remainder of this paragraph to

[19] We refer to [9] for a formal description of these ideas.

[20] http://kaon2.semanticweb.org/

[21] The same internet technology as the original Piggy Bank interface (Chrome, XUL, Javascript, XPCOM, Jetty and Ajax) is used.

[22] As WordNet mainly contains general vocabulary, more domain specific terminological sources will be needed. A suggestion found in the literature is to refer to Wikipedia entries.

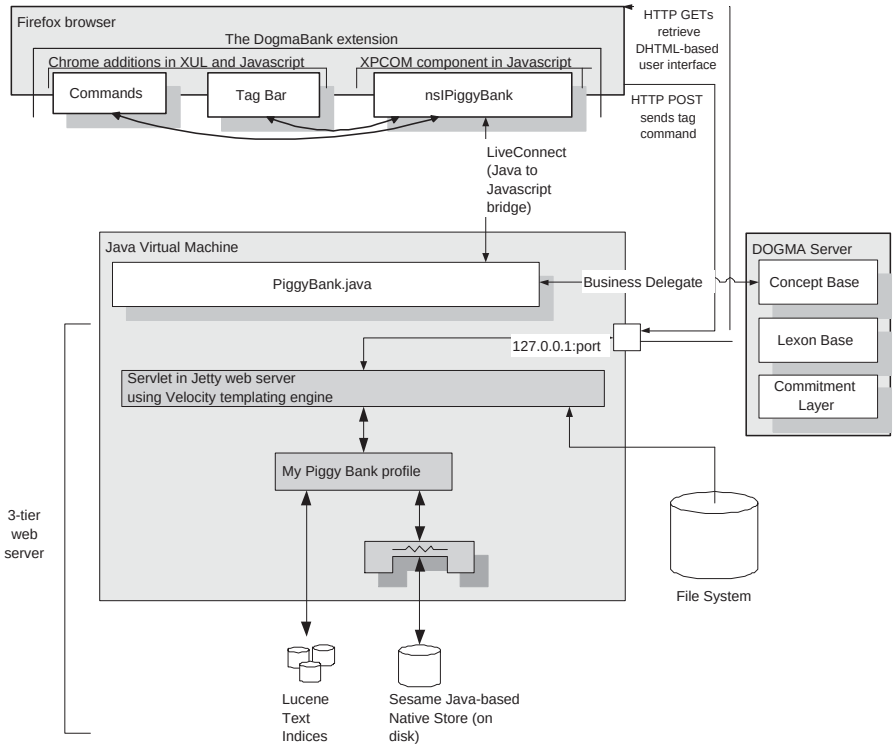

Fig. 1. The system architecture - partially adapted from [27]

Figure 2 to illustrate how one can tag web pages with concepts (step 1). The user types in a keyword (step 2), hits the fetch button or enter key (step 3) and a list of potential senses (i.e., informal lexicographic definitions of its different meanings or concepts) appears as well as some additional explanatory descriptions or glosses (e.g., examples of use). Retrieving senses is performed in two stages: (i) the synsets are checked on the presence of the term (e.g., 'car') and (ii) associations between a term and concept(s) as recorded in the DOGMA Ontology Server are looked up. Duplicates are removed and the resulting senses are displayed. The user now selects a particular sense (step 4) to tag the web page (step 5). It is also still possible to tag a web page with a *freely chosen keyword* (arrow 6). A user might be waiting for a concept to be added to the ontology by the moderator (see below), and decides to use such a keyword for the time being. Or he might want to mark a web page as "toread", which is a popular tag on current bookmarking tools, but without any relationship at all with the content or topic of the web page. All tags being alphanumeric, the original Piggy Bank system modules are re-used as they are (the lower part of Figure 1). Only, we have subverted (or rather 'smartened') the Piggy Bank system by having it store concept labels pointing to explicitly and commonly defined senses rather than personal interpretations of natural language words.

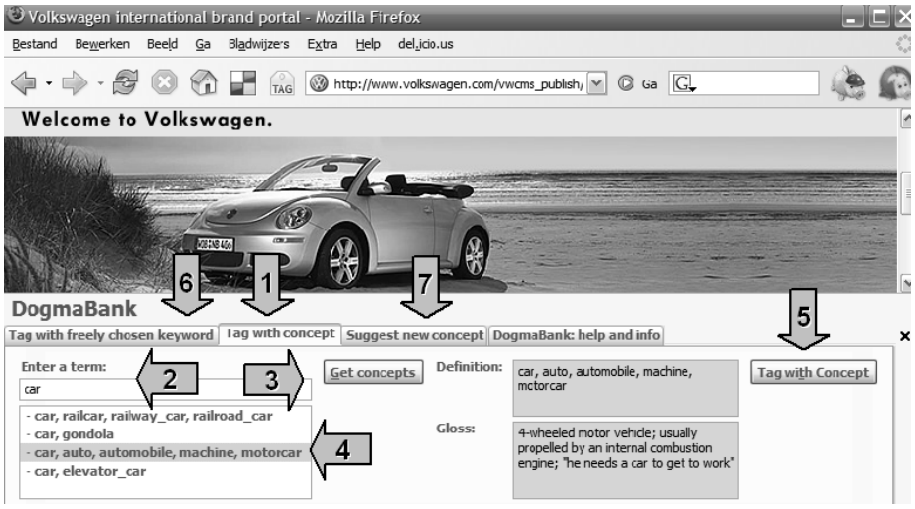

Fig. 2. The DogmaBank Firefox plug-in input and tag window with the senses for the term 'car'

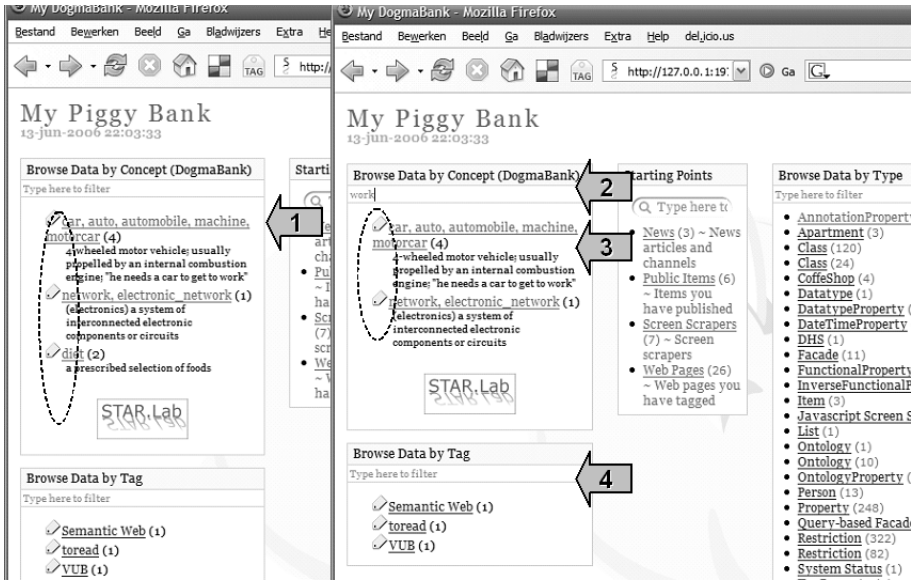

Fig. 3. The slightly adapted Piggy Bank interface showing a user's DOGMA concept tags and descriptions and illustrating the "refine as you type" functionality

The same mechanisms as used by folksonomies (usage frequencies, URLs per tag and tags per URLs) are now available to facilitate the meaning negotiation process for ontologies. It constitutes an interesting bottom-up complement (or maybe even alternative) to the traditional top-down or middle-out approach of engineering ontologies

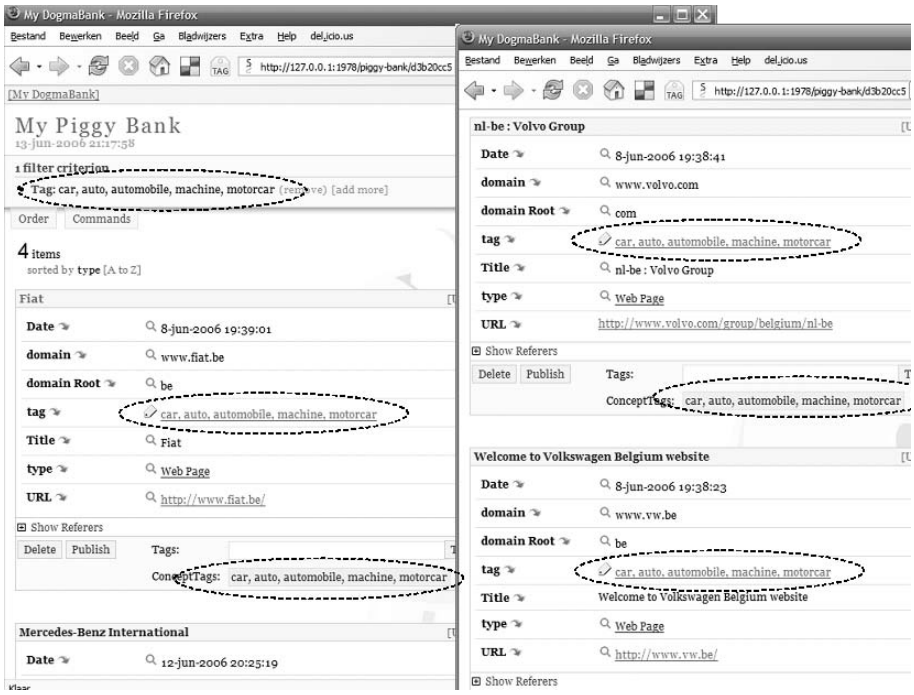

Fig. 4. The Piggy Bank interface showing the search results for the concept "car"

- see also [42]. Note that we are mainly considering what some call "lightweight ontologies", i.e. first order logic vocabularies without associated inference rules, but with explicit definitions for the vocabulary signature[23]. These mechanisms will provide the "ontology stakeholder committee" mentioned above with feedback on how their ontology is being used in practice. So one cannot create new concept tags at will, instead one suggests new tags with their definition (arrow 8 of Figure 2) to a committee of representative stakeholders for acceptance or rejection [24] to avoid an ontology from wildly and anarchically expanding. In order to avoid ending up with a huge collection of sligthly varying definitions, similarity checks have to be performed before a new concept description is entered in the system. Even the existing "folksabularies" are useful for ontology creation. In the same way as natural language processing methods cluster words of a corpus based on their distribution, tags can be "clustered" around the URL tagged. In both cases, it is assumed that the words or tags express somehow a similar meaning. Experience shows that this kind of information is very helpful when building an ontology. Formal definitions still have to be added by ontologists as this requires extra technical skills. The formal definitions mainly serve for machine-machine interoperability purposes, such as semantic web services. Also user-machine interaction

[23] It might be possible that inheritance and subclass relationships have been defined but are transparent for human taggers. More research needs to be done here.

[24] This part of the DogmaBank system is still under development.

(e.g., searches) will be improved by applying taxonomic and other reasoning mechanisms (see e.g. [21, 37]). For that purpose, formal relationships between the concepts are needed. In a first phase defining relationships are restricted to ontologists as this activity is more complex. Semantic Web tools [16, p.293 ff.] support this activity.

Also for information browsing and retrieval purposes - see Figure 3, a user will be able to enter keywords (step 1) and choose the appropriate sense as displayed by the system (step 3). To optimise the latter step, a "refine while you type" functionality has been implemented that searches the synsets and related term sets (step 2). It limits the number of potentially relevant senses shown to the user. The DogmaBank system is able cope with synonymy, homonymy and multilinguality - the latter not yet finalised. The Piggy Bank software has been slightly adapted to include the "browse data by concept" pane. Searching by key word tags is still possible (arrow 4). Figure 4 shows the search results for "car".

5 Related and Future Work

In this paper, we examined and illustrated how informal semantics resources (i.e. folksologies) and formal semantics resources (i.e. ontologies) meet productively in the form of an ESS supporting Internet-based communities in their collaborative work. Our analysis and upgrade of a current state-of-the-art system has helped us identify new focal points as well as gaps in current research and development: semantic matching and the development of similarity measures, and the groupwise usage of the information gathered via the (social) bookmarking process (including ways how to operationalise a stakeholder committee [11] and methods how to formalise a decision process leading to meaning consensus [4]) emerge as key issues here.

We will extend the DogmaBank system to transparantly determine at creation time whether or not the intended meaning of a new tag (i.e. its sense) is similar to an already existing one. Therefore, semantic similarity measures (e.g., [15, 33]) will be of prime importance not only for the work mentioned in this paper, but for many Semantic Web applications. Euzenat et al. [13] have gathered a large number of metrics of different origins that measure semantic distance (albeit in the context of ontology aligning and merging) between concepts. The key point will be to identify the most relevant and effective ones to be incorporated in DogmaBank-alike software.

Another area of related work concerns semantic and community portals (e.g., [12, 30, 32, 52]). A community portal stores and allows retrieval of all kinds of information of interest to a certain community. It is a centralised repository in which community members following a work flow upload data and annotate it with pre-defined tags, which might belong to an ontology, for easy retrieval purposes. A semantic portal basically adds semantic technology (i.e., searching on ontological meta-data instead of regular data) to retrieve information more accurately. The big difference with our upgraded social bookmarking tool is that a semantic community portal receives its meta-data from a central repository or from a crawler that gathers meta-data instances and URLs from all over the Web. In this case, the web site creators have to provide for the meta-data schema to be used by the crawler. No new tags can be added. In the context of the DogmaBank social bookmarking tool, a human more or less acts as the crawler (i.e.

encounters relevant information and decides to semantically tag a web page and store the reference). The "regular" Piggy Bank store however offers much less sophisticated retrieval possibilities (cf. Figure 4) than e.g., the OntoWeb semantic portal [44]. From the above, it should be clear that a semantic community portal is the natural complement of a semantically upgraded social bookmarking tool.

Also, many semantic annotating tools exist - cf. [16, p.344 ff.] for an overview [25]. The main difference with DogmaBank is that these tools only load a fixed set of tags of a predefined (and visualised) ontology. Tagging is an individual activity and happens with a finer granularity (single information items (instances) rather than entire web pages). When combined with language technology, it mostly concerns named entity recognition - cf. e.g., [35, 46]. These tools can also be used for automated ontology population.

Semantic annotators, crawlers, social bookmarking tools and community portals are applied in different contexts and settings and complement each other. They all contribute to the same overall goal, i.e. adding meta-data to the web and offering a means to share within a community information on the web. The challenge in ESSs is to find bridges between these automated and human approaches to creating and using formal semantics, thus working towards hybrid, semi-automated approaches, which is exactly what our DogmaBank tool aims at.

6 Conclusion

Semantics play an ever more important role on the Internet. Two major streams in research and development are distinguished: the world of formal semantics, exemplified by ontologies and the Semantic Web, and the world of informal semantics, of which folksologies (as frequently used by the Web2.0 community) are important bearers. The mission of the Semantic Web is to promote interoperability of information resources, so that machine agents become better at information retrieval. The Web2.0 wants to exploit the power of human communities to do the same.

Increasingly, the two worlds meet. In order to tap the full potential of the Web, emergent semantics systems will be required: well-designed socio-technical systems of formal and informal semantics, filled and used by well-calibrated combinations of machine and human agents. To help to analyse the larger picture (including the full range of semantics' functionalities in their context of use [48], which paves the way to a Pragmatic Web [47, 53], there is a need for the development of a meta-model of emergent semantics systems. We have implicitly already presented some elements of such a meta-model.

The main message of this paper is that there is a need and possibility (proven by our DogmaBank implementation, albeit still a preliminary one) to shift from thinking in terms of individual techniques to more holistic systems supporting communities of practice. In that way, it becomes much easier to identify the linkages and gaps between these techniques, and opportunities for new applications. We hope that this paper will contribute to further catalyzing and focusing the fundamental emergent semantics debate which is currently defining the future of a World Wide Web, be it as a Web2.0, a Semantic or Pragmatic web, or a combination of all.

[25] See also http://annotation.semanticweb.org/tools/.

Acknowledgements

Part of this research has been financed by the Flemish OntoBasis project (IWT GBOU 2001 #10069) [PS] and the DIP project (EU IST - FP6 507483). With many thanks to Luk Vervenne for pointing us to Piggy Bank.

References

1. Karl Aberer, Tiziana Catarci, and Philippe Cudré-Mauroux et al. Emergent semantics systems. In M. Bouzeghoub, C. Goble, V. Kashyap, and S. Spaccapietra, editors, *Semantics for Grid Databases, First International IFIP Conference on Semantics of a Networked World: ICSNW 2004, Revised Selected Papers*, volume 3226 of *LNCS*, pages 14–43. Springer, 2004.
2. Karl Aberer, Philippe Cudré-Mauroux, and Manfred Hauswirth. A framework for semantic gossiping. *SIGMOD Record Special Issue*, 31(4):48 – 53, 2002.
3. C. Batini, M. Lenzerini, and S. Navathe. A comparative analysis of methodologies for database schema integration. *ACM Computing Surveys*, 18(4):323–364, 1986.
4. C. Behrens and V. Kashyap. The "Emergent" Semantic Web: A consensus approach for deriving Semantic Knowledge on the Web. *Real World Semantic Web Applications*. Frontiers in Artificial Intelligence and Applications, Vol 92, IOS Press, 2002.
5. T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 284(5):34–43, 2001.
6. Andrew Burton-Jones, Veda Storey, and Vijayan Sugumaran. A semiotic metrics suite for assessing the quality of ontologies. *Data and Knowledge Engineering*, 55(1):84 – 102, 2005.
7. S. Buvac. Resolving lexical ambiguity using a formal theory of context. In Van Deemter and Peters, editors, *Semantic Ambiguity and Underspecification*. CSLI Publications, 1996.
8. Philippe Cudré-Mauroux and Karl Aberer et al. Viewpoints on emergent semantics. *Journal of Data Semantics*, 2880. to appear.
9. P. De Leenheer and R. Meersman. Towards a formal foundation of DOGMA ontology Part I: Lexon base and concept definition server. Technical Report STAR-2005-06, STAR Lab, Brussel, 2005.
10. Aldo de Moor. Patterns for the pragmatic web. In *Proc. of the 13th International Conference on Conceptual Structures (ICCS 2005)*, volume 3596 of *LNAI*, pages 1–18. Springer, 2005.
11. Aldo de Moor, Pieter De Leenheer, and Robert Meersman. DOGMA-MESS: A meaning evolution support system for interorganizational ontology engineering. In *Proc. of the 14th International Conference on Conceptual Structures (ICCS 2006)*, LNCS. Springer, 2006.
12. L. Ding, T. Finin, A. Joshi, Y. Peng, R. Scott Cost, J. Sachs, R. Pan R, P. Reddivari, and V. Doshi. Swoogle: A semantic web search and metadata engine. In *Proceedings of the 13th ACM Conference on Information and Knowledge Management*. ACM, 2004.
13. J. Euzenat, T. Le Bach, J. Barrasa, P. Bouquet, and J. De Bo et al. State of the art on ontology alignment. Knowledge Web Deliverable #D2.2.3, INRIA, Saint Ismier, 2004.
14. James Farrugia. Model-theoretic semantics for the web. In *Proceedings of the WWW2003 Conference*, pages 277 – 287, 2003. http://www2003.org/cdrom/.
15. F. Giunchiglia and M. Yatskevich. Efficient semantic matching. In Gomez perez A. and Euzenat J., editors, *Proceedings of the 2nd European semantic web conference (ESWC'05)*, volume 3532 of *Lecture Notes in Computer Science*, 2005.
16. Asunción Gómez-Pérez, Mariano Fernández-López, and Oscar Corcho. *Ontological Engineering*. Springer Verlag, 2004.
17. W. Grosky, D. Sreenath, and F. Fotouhi. Emergent semantics and the multimedia semantic web. *SIGMOD Record Special Issue*, 31(4):54 – 58, 2002.

18. T. R. Gruber. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 6(2):199–221, 1993.
19. Tom Gruber. Ontology of folksonomy: A mash-up of apples and oranges. In *First On-line Conference on Meta-Data and Semantics Research (MTSR05)*, 2005. http://tomgruber.org/.
20. N. Guarino and P. Giaretta. Ontologies and knowledge bases: Towards a terminological clarification. In N. Mars, editor, *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*, pages 25 – 32, Amsterdam, 1995. IOS Press.
21. N. Guarino, C. Masolo, and G. Vetere. Ontoseek: Content-based access to the web. *IEEE Intelligent Systems*, May-June4-5:70–80, 1999.
22. Nicola Guarino. Formal ontologies and information systems. In Nicola Guarino, editor, *Proceedings of FOIS '98*, pages 3 – 15. IOS Press, 1998.
23. Terry Halpin. *Information Modeling and Relational Databases*. Academic Press, 2001.
24. Jens Hartmann, Peter Spyns, Diane Maynard, Roberta Cuel, Mari Carmen Suarez de Figueroa, and York Sure. Methods for ontology evaluation. KnowledgeWeb Deliverable #D1.2.3, 2005.
25. Dion Hinchcliffe. 10 issues facing web2.0 today. http://web2.wsj2.com/10_issues_facing_web_20_going_into_2006.htm#, 2005.
26. Andreas Hotho, Robert Jäschke, Christoph Schmitz, and Gerd Stumme. Information retrieval in folksonomies: Search and ranking. In *Proceedings of the 3rd European Semantic Web Conference*, LNCS. Springer, 2006.
27. David Huynh, Stefano Mazzocchi, and David Karger. Piggy bank: Experience the semantic web inside your web browser. In *Proceedings of the third International Semantic Web Conference (ISWC05)*, LNCS. Springer Verlag, 2005.
28. M. Jarke, M. Lenzerini, Y. Vassiliou, and Y. Vassiliadis. *Fundamentals of Data Warehouses*. Springer-Verlag, 1999.
29. M. Jarrar and R. Meersman. Formal ontology engineering in the dogma approach. In R. Meersman, Z. Tari, and al., editors, *On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE*, volume 2519 of *LNCS*, pages 1238 – 1254. Springer, 2002.
30. H. Lausen, Y. Ding, M. Stollberg, D. Fensel, R. Lara, , and S.K. Han. Semantic web portals: state-of-the-art survey. *Journal of Knowledge Management*, 9(5), 2005.
31. Ben Lund, Tony Hammond Martin Flack, and Timo Hannay. Social bookmarking tools (ii): a case study - Connotea. *D-Lib Magazine*, 11(4), 2005. http://www.dlib.org/dlib/april05/lund/04lund.html.
32. A. Maedche, S. Staab, R. Studer, Y. Sure, and R. Volz. SEAL — tying up information integration and web site management by ontologies. *IEEE Data Engineering Bulletin*, 25(1):10–17, March 2002.
33. Ana Maguitman, Filippo Menczer, Heather Roinestad, and Alessandro Vespignani. Algorithmic detection of semantic similarity. In *Proc. of the WWW2005 Conference*. ACM, 2005.
34. Adam Mathes. Folksonomies: Cooperative classification and communication through shared metadata. http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html, 2004.
35. Diane Maynard. Benchmarking ontology-based annotation tools for the semantic web. In *UK e-Science Programme All Hands Meeting (AHM2005) Workshop Text Mining, e-Research and Grid-enabled Language Technology*, 2005.
36. Stefano Mazzocchi. Folksologies: de-idealizing ontologies. Blog http://www.betaversion.org/~stefano/linotype/news/85/, 2005.
37. Deborah McGuinness. Question answering on the semantic web. *IEEE Intelligent Systems*, (january/february):82 – 85, 2004.
38. R. Meersman. Semantic web and ontologies: Playtime or business at the last frontier in computing ? In *NSF-EU Workshop on Database and Information Systems Research for Semantic Web and Enterprises*, pages 61 – 67. NSF-EU, 2002.

39. Robert Meersman. The use of lexicons and other computer-linguistic tools in semantics, design and cooperation of database systems. In Y. Zhang, M. Rusinkiewicz, and Y. Kambayashi, editors, *The Proceedings of the Second International Symposium on Cooperative Database Systems for Advanced Applications (CODAS99)*, pages 1–14. Springer, 1999.
40. Peter Merholz. Metadata for the masses. http://www.adaptivepath.com/publications/essays/archives/000361.php, 2004.
41. Peter Merholz. Clay Shirky's viewpoints are overrated. http://www.peterme.com/archives/000558.html, 2005.
42. P. Mika. Ontologies are us: A unified model of social networks and semantics. In *Proc. of the 4th Internat. Semantic Web Conf. (ISWC05)*, LNCS, pages 522–536. Springer, 2005.
43. S. Nirenburg and V. Raskin. Ontological semantics, formal ontology, and ambiguity. In *Proceedings of the Second International Conference on Formal Ontology in Information Systems*, pages 151 – 161. ACM Press, 2001.
44. Daniel Oberle and Peter Spyns. *Handbook on Ontologies*, chapter OntoWeb - Knowledge Portal, pages 499 – 516. International Handbooks on Information Systems. Springer, 2004.
45. C.K. Ogden and I.A. Richards. *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. Routledge & Kegan Paul Ltd., London, 10 edition, 1923.
46. Borislav Popov, Atanas Kiryakov, Damyan Ognyanoff, Dimitar Manov, and Angel Kirilov. KIM - a semantic platform for information extaction and retrieval. *Journal of Natural Language Engineering*, 10(3-4):375 – 392, 2004.
47. Mareike Schoop, Aldo de Moor, and Jan Dietz. The pragmatic web: a manifesto. *Commun. ACM*, 49(5):75–76, 2006.
48. Amit P. Sheth, Cartic Ramakrishnan, and Christopher Thomas. Semantics for the Semantic Web: "The Implicit, the Formal and the Powerful". *International Journal on Semantic Web and Information Systems* 1(1): 1–18, 2005.
49. Clay Shirky. The semantic web, syllogism, and worldview. http://www.shirky.com, 2003.
50. Clay Shirky. Ontology is overrated: Categories, links and tags. http://www.shirky.com, 2005.
51. Peter Spyns, Robert Meersman, and Mustafa Jarrar. Data modelling versus ontology engineering. *SIGMOD Record Special Issue*, 31 (4): 12 - 17, 2002.
52. P. Spyns, D. Oberle, R. Volz, J. Zheng, M. Jarrar, Y. Sure, R. Studer, and R. Meersman. Ontoweb - a semantic web community portal. In D. Karagiannis and U. Reimer, editors, *Proceedings of the Fourth International Conference on Practical Aspects of Knowledge Management (PAKM02)*, volume 2569 of *LNAI*, pages 189–200. Springer Verlag, 2002.
53. Peter Spyns and Robert Meersman. From knowledge to interaction: from the semantic to the pragmatic web. Technical Report 05, STAR Lab, Brussel, 2003.
54. P. Spyns and J. De Bo. Ontologies: a revamped cross-disciplinary buzzword or a truly promising interdisciplinary research topic? *Linguistica Antverpiensia NS*, (3):279–92, 2004.
55. P. Spyns and M.-L. Reinberger. Lexically evaluating ontology triples automatically generated from text. In A. Gómez-Pérez and J. Euzenat, editors, *Proceedings of the second European Semantic Web Conference*, volume 3532 of *LNCS*, pages 563 – 577. Springer Verlag, 2005.
56. Peter Spyns. Object Role Modelling for Ontology Engineering in the DOGMA framework. in R. Meersman and Z. Tari and P. Herrero et al. *On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops*, volume 3762 of *LNCS*, pages 710 – 719. Springer Verlag, 2005.
57. M. Uschold and M. Gruninger. Ontologies: Principles, methods and applications. *Knowledge Sharing and Review*, 11(2), June 1996.
58. M. Ushold and M. Gruninger. Ontologies: Principles, methods and applications. *The Knowledge Engineering Review*, 11(2):93 – 155, 1996.