Can the Pragmatic Web Assist Search Engine Optimisation?

Tiana Razafimbelo[1], Simon Polovina[2], Richard Hill[2]
[1]Eleventeenth Limited, Wakefield, UK
[2]Sheffield Hallam University, Sheffield, UK
tiana@eleventeenth.com
{s.polovina, r.hill}@shu.ac.uk

Abstract: Discovering information on the Web relies on the publishers of that information to describe it in a meaningful way, often at the syntactical level. Despite ongoing efforts to improve Web page description, search retrieval is more time consuming than it should be. Essentially this arises because search engines lack the capacity to identify the purpose of the search. The goal of this paper is to identify the extent to which this issue can be addressed and how the Pragmatic Web may assist in this endeavour. The paper uses Stamper's Semiotic Ladder to distinguish between those issues that are truly semantic, pragmatic or even social, versus those which are syntactic. This enables the core issues to be discovered, thereby distinguishing the areas that the Pragmatic Web community can focus upon.

1 Introduction

Search Engine Optimisation (SEO) is the process of "preparing a website to enhance its chances of being ranked in the top results of a search engine once *a relevant search* is undertaken" [WW06] (our italics).

Ranking techniques differ from one search engine to another, mostly depending upon the content of the web pages published to the web [BN99]. Content encapsulated in Hypertext Mark-up Language (HTML) tags is typically extracted and then used to rank the pages, displaying information to the user, such as the information in the HTML title tag. Similarly, the HTML Meta tag for the description would be used to display the first few sentences of the page in the search results.

This means that should the information be incorrectly represented at the syntactic level, it could impair the indexing and therefore the search results [P03].

To be indexed, web pages are analysed by software programs called crawlers. The data gathered from web pages is filtered using mechanisms such as text normalisation and segmentation, word weighting, word stopping and word stemming. It is then saved in the search engine's 'database' to index the pages accordingly [BN99]. This data is then used to calculate the 'relevance score' of the web pages, which in turn is used to rank the document in relation to other documents [D03]. Thus pages with high relevance scores would be the first ones to be displayed in the search results.
However, even when the content of the pages is adequate, it can be difficult to get useful results from search engines: there is an issue of accuracy, relevance and usefulness for the results fetched by search engines. This is exacerbated by the increased amount of material on the Web, as well as lack of uniformity of document structure [P03]. The proliferation of heterogeneous data sources compounds this problem, leading to a loss of search engine efficiency and the delivery of erratic results, especially when the query is vague.

Due to this inefficiency, numerous companies have specialised in delivering SEO services assuring high ranking for major search engines such as Google or Yahoo. Those companies are using methods such as 'keywords stuffing' – altering the natural content of the page to add a 'bunch' of keywords that may or may not have a direct link with what the website or web page is about, usually altering the natural meaning of a page. The purpose of using those methods is to allow sites to be highly ranked. Tampering with the web pages such as described may lead to some sites being ranked highly for certain keywords, even though they do not necessarily deal with the subject surrounding those keywords.

In this paper we explore how pragmatics along with semantics can assist in SEO. We seek to reorient the work undertaken by web site owners back to their meaningful content, thereby better fulfilling the user's search experience. We concentrate our efforts on the issues described as follows:

- How data contained in the web pages can be used to allow more efficient searches;
- How can we make sure that the data analysed in the page is contextually the one the user expects/wants;
- How can we make sure that even if the data belongs to the correct 'context', the content was not tampered with in a bid to highly rank the pages.

2 Information Retrieval and the Semiotic ladder

To assist our investigations we refer to Stamper's semiotic ladder (Figure 1) to provide some potential classifications.

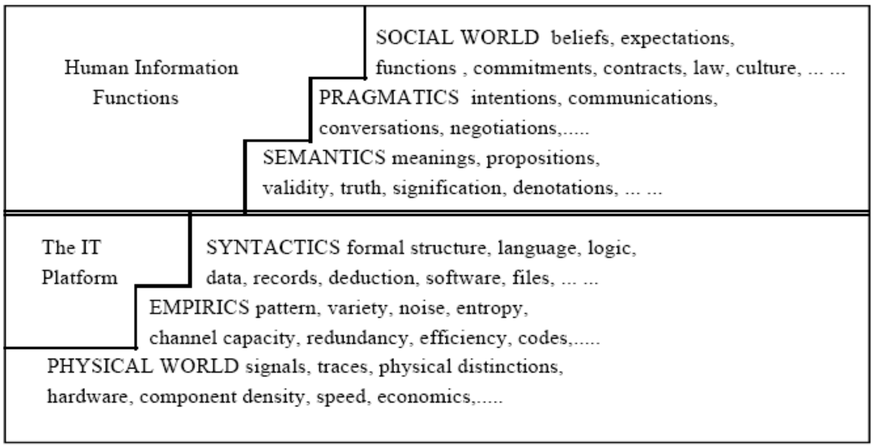

Figure 1: Stamper's Semiotic Ladder [S96]

Stamper's model is appropriate since it is widely understood and "the refinement offered by the ladder can be used to investigate a certain layer in more detail" [MW02]. It also helps examine all the aspects involved for communication and coordination in an organizational context [SL02]. The semiotic ladder consists of the views or signs from the perspective of physics, empirics, syntactic, semantic, pragmatics, and the social world. The additional social world layer stresses that information use is always a part of human behaviour in a social setting, whereas norms or social conventions govern people's behaviour ([L00]).

2.1 The Syntactic and Semantic Layers

The two steps of the ladder that have attracted the most attention in the past decade are the syntactic and semantic 'rungs'. Figure 1 illustrates how the Syntactic web can be viewed as a set of documents and subsequently how their content such as elements, links and tags can be represented. The semantic web/layer represents how data can be extracted from web pages using techniques such as XML, RDF and used to create ontologies. The layers thus provide a lucid framework to understand the implementations of standards and to allow data to be extracted and understood.

2.2 An Experiment with the Pragmatic Layer

The pragmatic layer concentrates on the context of information. To illustrate the fact that context identification is crucial for search engine optimisation, we performed two searches using the Google.co.uk search engine.

Firstly, ‘pragmatic’ and ‘web’ were entered as search terms. Our goal was to find the www.pragmaticweb.info website, in order to retrieve the exact URL of this site. The ensuing results illustrated that the site was not listed under this URL on the first 10 pages, but was appearing on Page 2 of the search results under an alias URL. This apparent obfuscation was confusing and suggests that users might not recognise the site at all (since the URL was not the one they expected to see). Interestingly the results included a website called `proxy.org'. This site is “the pragmatic Web surfer's guide” and leads with privacy issues – such as cookies - and proxy settings when browsing sites. The words *pragmatic* and *web* are within the homepage but their meaning is *different* than the one intended when performing the search. In this particular case, the word *pragmati*c refers to the surfer’s guide *not* the web and the word *web* is the first word of the expression ‘web surfing’. This site does not deal with the Pragmatic Web but is a practical/pragmatic guide to web surfing. This illustrates that the result set did not take into account the meaning in which the search was performed and therefore we conclude that the context of the search was ignored.

As a comparison another search performed using the terms ‘*Search Engine Optimisation*’ retrieved results about companies offering Search Engine Optimisation as a service as opposed to articles or web pages discussing SEO techniques that be used within web pages.

2.3 The Social Layer

This later example showed that the number of pages associated with particular keywords dictate what results will be presented to the user; clearly the amount of data influences the results of the search. Since it is the quality of the data as well as the quantity of the data that is taken into account, making sure that the data used is of the correct context *and* of the trusted source is a major issue. This has even been raised as an issue by the Google executive director, who emphasised the need to verify the source of the web page to allow the content to be captured or not [L06].

With this issue of trust, we need to take into account the social factors behind a search engine: how one search engine ranks and values a site, can have tremendous effects on site visibility (for example boosted sales in the case of an e-commerce site). How this site is perceived and ranked in search engines allows browsers to assess its social popularity and notoriety, which will have an impact on a user’s decision to visit the site or not. This can be represented in Stamper's ladder as the Social layer - the perception/belief of ones when placed in a site. It shows the social aspects linked to the success of SEO (or not).

Without acknowledging and tackling this issue of trust, tampered content will still be part of the search results. This issue of trusts means that efforts need to be concentrated on both pragmatic and social layers for this issue to be resolved.

There is a need for a robust process or procedure in order to allow this trust to take place. A common example of this issue is when looking for a footwear website in the hope of buying online, and finding a site that supposedly is about shoes but in fact sells something totally different. This is an issue to be addressed as part of the social layer.

2.4 Summary of findings

The table below summarises the main two aspects we have outlined in the previous sections: not only can the pragmatic web assist in optimising search engines but the social layer will play a significant role too.

Layer Name	**Aspects within this layer**
Social layer	Trust of source of information (identification of source) social ranking – popularity Social Impact on the identity of the site AND on the users to visit it
Pragmatic layer	Usage determination of the context of the search
Semantic layer	Meaning using Standards such as XML, RDF or OWL / the tags

Table 1: Adapted Semiotic ladder from a Search Engine Optimisation perspective

3 Realising Context Identification in Search Engines

We have seen in previous sections that the role of the pragmatic web would be central in identifying the context of the search and therefore in making search results more relevant. We now want to present an idea of how this could be realised.

The semantic web is an extension of the current Web in which information is given a "well-defined meaning, enabling computers and people to work in better cooperation" [P03]. In principle, it makes it possible to retrieve information from the web using extensible mark-up techniques based upon XML, such as RDF, to structure documents, using concepts, properties and values to express meaning. More importantly, this facilitates the generation of ontologies that describe concepts and relations between them [M05].

Sowa states that "*ontology is the study of the categories of things that exist or may exist in some domain. The product of such a study, called an ontology, is a catalogue of the types of things that are assumed to exist in a domain of interest D from the perspective of a person who uses a language L for the purpose of talking about D.*" [S99]. Ontologies are crucial for the classification of concepts, and become more powerful when combined with rules that determine how the ontology is processed.

The hypothesis is that if the ontology is correctly identified, it will mean that the context of the search has been either identified correctly (100% match) or the ambiguity has been decreased. Then the search results are bound to be more relevant. The ontologies retrieved can then be used to filter the web documents returned by the search engine when performing the search on web documents.

3.1 Structuring Ontologies

We propose that searches use ontologies as a means to identify the context of use. We envisage that this step would be encapsulated from the user by default, as it is desirable to use the same user-friendly mechanism to search for documents as to search for ontologies.

Accordingly, each document is represented as a vector of *n* coordinates, where each coordinate corresponds to a keyword. The dimension *n* of the space in which the vectors are created is the total number of documents. Coordinates reflect the positions of documents with each other. Similarities between documents are therefore represented by their coordinates and similarities with a query expression are simply calculated using the cosine between the query and each document in turn [W03a].

Figure 2 shows how in a 2-dimension space a search vector can be represented against a few terms already existing within this space.

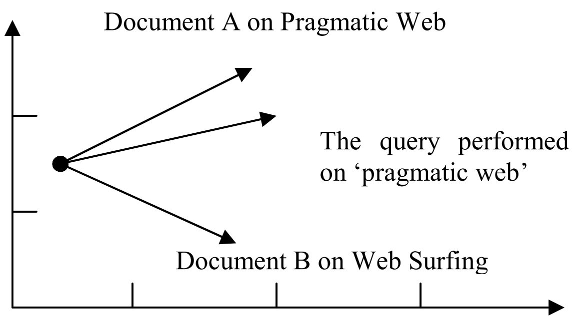

Figure 2: Vector representation of search terms for query 'pragmatic web'

Visually we can see how close document A is from the query. Similarities are calculated using algorithms. In this definition, vectors bear similarities to line diagrams concept lattices from Formal Concept Analysis (FCA):

- documents are the objects and the shared coordinates form all the attributes that a document may contain;
- the representation of objects map the way vectors are viewed;
- one object can be found using several routes and similarly several keywords, which is essential in the search engine perspective that we are taking;
- concept lattices have a powerful `switching scale' concept that allows documents to be displayed differently depending on the 'view'/filter chosen. This is related to work developed on Mail Sleuth [DE05] but also FCA's use in browsing documents [DMS05], [P04], [GM03].

3.1.1 Concept Lattices for ontologies

FCA is a method described by Rudolf Wille, whose main representation is called concept lattices. Concept lattices are used to represent the interaction between objects and their attributes [P06]. An example of an object is a cat and an attribute is its colour e.g. red. The concept lattices are a core part of our proposed solution to SEO.

Fundamental to concept lattices is the concept of duality called "Galois connection". This describes the relation between objects and attributes and how objects are connected between them by one or more attributes [P06]. One such example is that an object is a document, and an attribute is one of the terms that are used to describe this document e.g.: a keyword. As an object can have more than one attributes, this document could potentially have associated with it a number of keywords, or terms. Concept lattices allow the representation of those objects to differ depending on what scale is used to model the diagram [P06].

A ‘scale’ in this context is similar to a query of search terms and applying it means that the results presented in the lattice will be *filtered* depending on the scale chosen: it is similar to displaying results according to particular search terms.
Scales can be nominative scales, where the attributes are names or an interval, and therefore the attributes can be used to build ranges such as age range or price range.

3.1.2 An Exemplar

In this section we show how we see the process working behind the scenes. Throughout the section, we will use the example stated in Section 2 regarding the search terms ‘pragmatic’ and ‘web’. Intermediate steps are shown in Figures 3, 4 and 5.

To each ontology is associated one or more keywords and one or more domain. Each domain can belong to one parent domain. This would be similar to a view of a taxonomy or simple tree structure where domains such as ‘the web’ include ‘pragmatic web’ and ‘semantic web’ for example.

The relationships between ontologies and keywords would be represented as a concept lattice as it allows keywords to belong to one or more ontologies as well as ontologies to be associated to one or more keywords. Likewise the relationships between web documents and keywords can again be a concept lattice because it allows more complex relationships to be represented.

If we were to generate a diagram to show the relationship between documents and ontologies based on the two previous structures, it will have to show that ontologies are based upon one or several documents. However one document’s domain can include one-to-many ontologies. As a result, a concept lattice would be the preferred choice simply because it offers this flexibility.

Let L be our lattice containing the representation of ontologies On.
The terms used for the search are ‘pragmatic’ and ‘web’.

Let O1 be the ontology representing the domain of knowledge ‘pragmatic web’
Let O2 be the ontology representing the domain ‘web research’
Let O3 be the ontology representing the domain ‘web surfing’
Let O4 be the ontology representing ‘the semantic web’.

Note that in our example, we are simplifying those concepts. The representation of those ontologies can be as follows:

O1 { Pragmatic web } Extends O4
Description keywords {pragmatic, web, meaning, concept, context}
O2 { Web research } is part of O1 and O4
Keywords { conference, research … }
O3 { web networking } extends 'computer network' ontology
Keywords { security, cookie, proxy }
O4 { Semantic Web }

From this classification, we can represent the ontologies in the lattices below:

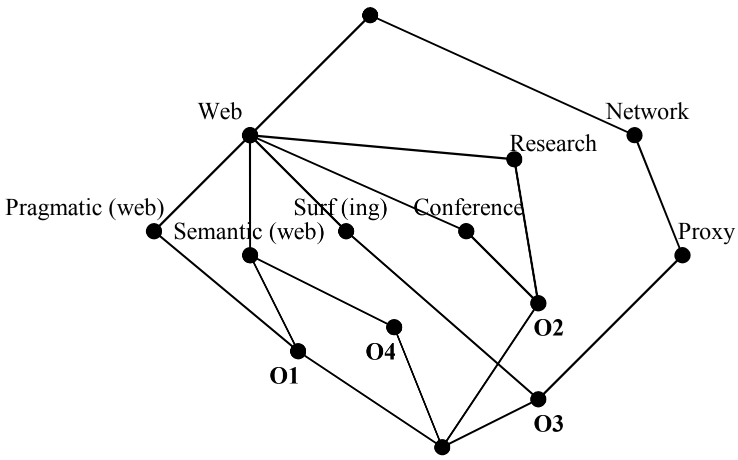

Figure 3: Simplified concept lattice representing relations between terms and ontologies

In our simplified example, if the user enters *pragmatic* and *web*, then those terms can be used to establish the correct meaning/context of use and subsequently narrow down the domains to look at. Figure 4 is presenting the up-to-date concept lattice after a nominal scale has taken place.

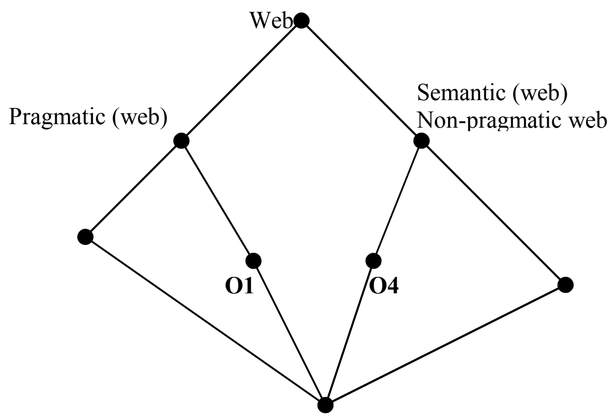

Figure 4: A nominal scale for the 'pragmatic web' based on Figure 3

Figure 4 shows that the context specified by the query can be represented by the ontology O1.

After identifying which ontology was best to use to answer this query, the correct ontology can be used to filter the results found by the search engines. Any ontologies/domains found can either be presented to the user (to allow them to choose the appropriate one if listed), or the results can be grouped by domain and displayed to the user.

Then the representation of the relationships between documents and ontologies can be automatically generated and shown in a concept lattice between ontologies and web documents. More explicitly, it could be established and represented in a topic map as shown in Figure 5 [P03]. Topic maps are used in this instance as an intuitive way to represent relations between topics/subjects and web documents. This is because the topics' hierarchy can be clearly represented and the idea of broadening the results would then mean including documents whose topics are also the 'parent topic'. Figure 5 shows only a snippet of a more global topic map to serve for our example shown by Figure 4. The topic information comes directly from the ontologies.

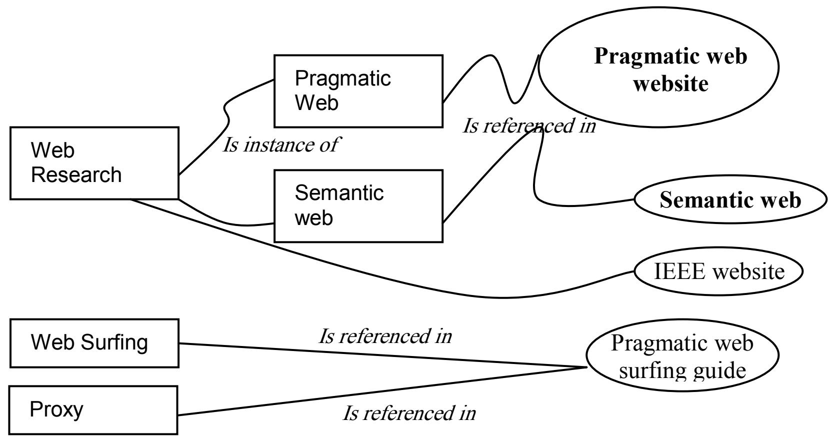

Figure 5: Example of relationships between web documents (by websites) and domains/ontologies

The 'is referenced in' links represent the fact that the website is about this particular topic. The 'is part of' relation in Figure 5 does not represent an include statement relationship. This means that those ontologies can be used to filter the search results, and positioned the web pages/documents whose information belongs to the same domain as the ontologies return at the top of the search results. This feature is powerful as also demonstrated by Eklund [DE05] as it allows representation of the objects within the lattice to be dynamically updated depending on this scale.

The relationships between ontologies can allow results to be broadened if needed. For example, as O1 extends O4, results could be enlarged by also including results that correspond to this ontology (in our example Figure 5, a website on the semantic web would be displayed). This can help in case the context was too precise and narrowed the searched unduly. In the same way, association relationships between ontologies can be used to display more results where meaning is not assured.

In Figure 5, we can see that selection of the appropriate ontology will rule out the "web surfer's guide" website from appearing in the search results. Also we know, O2 is part of O1, therefore all knowledge encapsulated in O2 could also be displayed (but either against a different context or further down the search results in terms of relevancy). In our example, the semantic web website and the IEEE website all refer to the ontology O2.

We have proposed a way to structure ontologies. Note that if the search terms entered do not permit the identification of one or more ontologies to focus on, we can propose to the user a series of domains based on the preliminary findings. Alternatively we can ask them to be more precise (offering guidance).

3.2 Publishing Ontologies

To allow the issue of trust to be tackled, there is a need for a structure in which ontologies will be added and controlled (by an authority such as W3C for example) in order to identify any source(s) and if they can be trusted. This implies the construction of a main ontology repository that would be used for search engines or any other online retrieval system. It will involve the user updating ontologies online and changes being checked (automatically or manually) to prevent from further tampering with data.

It may be interesting to put in place a registration process that users will have to go through in order to update or publish ontologies. This 'registration' will enable basic details to be fetched and allow the user to be 'certified' as a trusted party.
All those details should then be available at the semantic level where they can be retrieved and shared using techniques such as RDF.

All trusted parties will be able to update ontological content and as a result the web pages they own will be detected first by search engines. This trust would help build upon the new resources on the web, which is why it is important for the Internet community to be educated as it involves a lot of processing and collaboration from the users in terms of the input expected from them.

Linked to this repository, can be associated further mechanisms to allow a more automated process to take place to reduce information that has been tampered with or false. Such processes can be for example as explained in [WD06] could be the detection of cloak pages on the web. This can be an added layer of security that moves towards a more practical web trust model.

The publishing of ontologies would also facilitate:

- the ability for relations between ontologies and documents and also ontologies between themselves to be drawn.
- the automatic generation of those relationships using topic maps or concept lattices, as these graphs allow the required many-to-many representation (multiple ontologies for one document, multiple documents for one ontology).

.

The information published by the user would allow relationships between ontologies to be known and generated as well as relations between documents and ontologies. This means that ontologies will ultimately complete one another but their relation would not be that of a parent-child, but more an association relationship. Those relations can be generated by submitting at least a set of minimum requirements as shown below in OWL:

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [
     <!ENTITY rdf  "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
     <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
     <!ENTITY xsd  "http://www.w3.org/2001/XMLSchema#" >
     <!ENTITY owl  "http://www.w3.org/2002/07/owl#" >
     <!ENTITY PragmaticWeb "A_URL" >
   ]>

<rdf:RDF
  xml:base  ="A_URL"
  xmlns     ="&pragmaticweb;"
  xmlns:pragmaticweb ="&pragmaticweb;"
  xmlns:owl ="&owl;"
  xmlns:rdf ="&rdf;"
  xmlns:rdfs="&rdfs;">

<owl:Ontology rdf:about="">
  <owl:versionInfo>$Revision: 1.0 $</owl:versionInfo>
</owl:Ontology>

<!-- Class Message represents a series of terms -->

<owl:Class rdf:ID="Message">
  <rdfs:label>Message</rdfs:label>
  <rdfs:comment>Message to identify meaning of</rdfs:comment>
  <rdfs:subClassOf rdf:resource="&owl;Thing" />
  <rdfs:subClassOf>
     <owl:Restriction>
          <owl:onProperty rdf:resource="#term" />
          <owl:minCardinality>1</owl:minCardinality>
     </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

<owl:DatatypeProperty rdf:ID="term">
  <rdfs:domain rdf:resource="#PragmaticWeb"/>
  <rdfs:range rdf:resource="&xsd;string" />
</owl:DatatypeProperty>

<!-- end of Class Message -->
</rdf:RDF>
```

3.3 User Involvement & Issues

Throughout this paper, both Stamper's social and pragmatic layers need attention in order to improve search engine optimisation. We see the social layer not as a layer itself but an extension of the pragmatic layer that will allow the latter to be 'successful', i.e. allowing context identification to take place. As discussed, publishing ontologies will involve the Internet community getting involved in the 'trusting' process. This means that:

- this task will be time consuming and therefore will demand automation;
- the measures put in place need a trust model that prevents tampering with the trusted data

As a whole, the user involvement in the overall process is crucial:

- The user will be documenting each new object that is to be part of the concept lattice. In a study [DE05] on a simple small intranet system such as MailSleuth, it was found that the user found this very constraining and time-consuming.
- It is conceivable that search engines will want to update the way the results are displayed by domain and then ranking as opposed to just overall ranking. This will allow the user to visually 'select' which context is the appropriate one and be a active actor of the search as opposed to a passive watcher.

It may be that not all the questions can be answered within the pragmatic web, but it will raise additional issues such as involvement of the user within the search and also within the publication of ontologies. Web services such as search engines may well have to adapt to accommodate this change and to encapsulate this user collaboration within the system.

There is clearly fertile ground for further investigation, in particular the:

- formulation of a trust model for ontology publishing;
- development of tools to facilitate the inter-operation between ontologies.

Such issues remain to be explored.

4 Conclusions

Search Engine Optimisation is a well-known marketing strategy to publicise a site onto the World Wide Web. However, it relies upon search engines finding the site and results are not always reliable. Using Stamper's Semiotic ladder, we illustrate that this issue could be dealt with at the *pragmatics level* with some implications for the social world.

We have addressed the issue of identifying the context of use and have proposed a means by which ontologies can be organised and browsed. We have also indicated that the user will directly influence the take-up of the Pragmatic Web, and thus, in the absence of automated tools, play a key role as to how the context can be identified.

Using the input of the user, the pragmatic web and semantic web will achieve improved optimization of information retrieval. However with the input of the user, the pragmatic web is also likely to turn towards the next layer, the *social world,* when dealing with user collaboration, trust and belief. As such, we believe that our novel approach combined with Stamper's Semiotic Ladder is a first step towards the development of a framework that describes how the Pragmatic Web can assist search engine optimisation.

References

[B02] T. Berners-Lee. *The Semantic Web lifts off.* 2002 http://www.ercim.org/publication/Ercim_News/enw51/berners-lee.html.

[BN99] R. Baeza-Yates, B. Ribiero-Neto. *Modern Information Retrieval.* 1999. Addisson Alley (Ed.) pp. 367-393.

[D03] D. Joubert. *Cross language spoken retrieval system using EuroWordNet.* 2003.

[DE05] S. Domingo, P. Eklund. *Evaluation of concept lattices in web-based mail browser.* 2005.

[DJ04] L. Ding et al. 2004. Swoogle: a search and metadata engine for the semantic web. In *Proceedings of the Thirteenth ACM international Conference on information and Knowledge Management* (Washington, D.C., USA, November 08 - 13, 2004). CIKM '04. ACM Press, New York, NY, 652-659. DOI= http://doi.acm.org/10.1145/1031171.1031289.

[DMS05] F. Dau, M-L. Mugnier, G. Stummer. *Contributions to ICCS 2005.* Kassel University Press. pp.54-65, pp.132-145, pp.172-185. 2005.

[GM03] B. Granter, A. de Moor. *Using conceptual structures – Contributions to ICCS 2003.* "Step-wise, Concept lattice navigation". 2003.

[KDFO03] M. Klein, Y. Ding, D. Fensel, B. Omelayenko. *Towards the Semantic Web: Ontology-driven Knowledge Management,* Ontology Management: Storing, Aligning and Maintaining ontologies 2003, pp. 47-69.

[L00] K. Liu. *Semiotics in information systems engineering.* Cambridge, England: Cambridge University Press (2000).

[L06] C. Lombardi. *Google exec challenges Berners-Lee* (18 July 2006) http://news.zdnet.com/2100-9588_22-6095705.html.

[M05] A. De Moor. Patterns of the Pragmatic Web. *Proc. of the 13^{th} International Conference on Conceptual Structures (ICCS 2005)* pp. 1-18, Kassel, Germany. 2005.

[MW02] A. De Moor, Weigand. Towards a Semiotic Communications Quality Model Published in Organizational Semiotics: Evolving a Science of Information Systems, Kluwer, Boston, 2002, pp.275-285.

[P03] T. Passin. *Explorer's guide to Semantic Web.* 2001. Chapter 5.

[P04] J. Pane. *Query Based Multicontext Based browsing: a technical report.* 2004.

[P06] U. Priss, *Formal Concept Analysis in Information Science.* In: Cronin, Blaise (ed.), Annual Review of Information Science and Technology. Vol 40, pp. 521-543. 2006.

[S96] R. Stamper. Signs, norms, and information systems. In *Signs at Work* (HOLMQVIST B *et al.*, Eds), Walter de Gruyter: Berlin, Germany. 1996, pp. 349–397.

[S02] M. Singh *The Pragmatic Web: Preliminary thoughts.* 2002.

[S99] J.F. Sowa. Knowledge Representation: Logical, Philosophical and Computational Foundations, Brooks-Cole, 2000.

[SL02] K. Liu, L. Sun *Information Systems Frontiers.* 2002. Vol 4, No. 3, pp 251-256. In the special issue on "Co-Design of business and IT Systems".

[SM03] P. Spyns, R. Meersman, *From knowledge to interaction: from the semantic to the pragmatic web.* 2003.

[T06] Tolk, A. 2006. What Comes After the Semantic Web - PADS Implications for the Dynamic Web. In *Proceedings of the 20th Workshop on Principles of Advanced and Distributed Simulation* (May 24 - 26, 2006). Workshop on Parallel and Distributed Simulation. IEEE Computer Society, Washington, DC, 55. http://dx.doi.org/10.1109/PADS.2006.39.

[W03a] D. Widdows, *Geometry & Meaning.* 2003: Chap5, Chap 8.

[W03b] D. Widows, Unsupervised methods for developing taxonomies by combining syntactic and statistical information. In *Human Langauge Technology / Conference of the North American Chapter of the Association for Computational Linguistics* (HLT/NAACL), Edmonton, Canada, pp. 276–283. 2003.

[W3C] W3C. *What is an ontology?* http://www.w3.org/TR/webont-req/#onto-def, 2006.

[WD06] B. Wu, B.D. Davison. Detecting semantic cloaking on the web. In *Proceedings of the 15th international Conference on World Wide Web* (Edinburgh, Scotland, May 23 - 26, 2006). WWW '06. ACM Press, New York, NY, 819-828. DOI= http://doi.acm.org/10.1145/1135777.1135901, 2006.

[WW06] SEO http://www.websearchworkshop.co.uk/glossary.php.