DCPAPERS
Proc. Int'l Conf. on Dublin Core and Metadata Applications 2012

Enriching Webpages with Semantic Information

Norbaitiah Ambiah
MIMOS BHD, Malaysia
norbaitiah.ambiah@mimos.my

Dickson Lukose
MIMOS BHD, Malaysia
dickson.lukose@mimos.my

Abstract

This paper proposes a tool to automatically enrich webpages with semantic information by annotating keywords in the document with microdata markup. There are two case studies described and implemented in this paper. The first case study focuses on generating new webpages with microdata and the second case study focuses on enriching existing webpages with microdata. This paper also demonstrates the practicality of using schema.org terms in constructing a referenced ontology. Finally, a comparative study is conducted and the result shows that the proposed tool is more reliable in terms of performance and advanced features compared to other existing automatic microdata generator tools.

Keywords: microdata; Schema.org; schema.rdfs.org; keyword extraction; semantic annotation; semantic webpage; html5.

1. Introduction

In the current practice, most of the conventional webpages display information in HTML markup tags where this information is usually stored in a structured repository (e.g. database or knowledge base). The HTML tags in the webpages remove the structure of the data which actually reflects the important meaning of the represented data. Due to this limitation, search engines such as Bing, Google and Yahoo! have to rely only on a keyword-based approach during the searching process. This approach, however, is subject to some significant drawbacks. The search results are not 100% accurate and may not be relevant to the user's actual request. Therefore, the effort to enrich webpages with semantic information is highly needed for search engines to not only return more accurate results but also return meaningful information.

Some notable efforts in enriching web pages with semantic information are that of Dobrev and Strupchanska (2005) which perform semantic annotation of web pages in a semi-automated manner, and the works of Kudelka et al. (2006, 2009) that generates semantic annotation of web pages using web patterns. In June 2011, Google, Bing, and Yahoo! introduced Schema.org, a collection of terms for structured data markup on web pages to improve the display of search results (Hausenblas and Cyganiak, 2011). The latest HTML (i.e., HTML5) is an impressive version where it improves the HTML presentational elements to have a new lightweight semantic meta-syntax called microdata. Microdata attributes are defined in nestable groups of name-value pairs of data and together with Schema.org terms, give a new way to add semantic information in the webpages (Studholme, 2011). The structure of semantic information embedded in a webpage gives an important meaning to the data it represents. Moreover, this metadata can be described logically in a graph notation and can enrich the data by linking the resources to Linked Open Data (LOD).

In this paper, we propose a tool called Schema.org Microdata Creator (ScheMicCr) for automatically generating microdata syntax to build semantically annotated webpages. Our main objective is to enrich webpage with semantic information so that it easier for search engines like Google, Bing and Yahoo! to create semantic indexes. ScheMicCr tool is designed to be flexible so it can be applied in any domain. However, in this paper, the experiment focuses on constructing semantic annotated webpages of patent information.

https://doi.org/10.23106/dcmi.952136024

This paper is structured as follows. In Section 2, we explain the methodology and the sample result of the experimental work. In Section 3, we highlight the result of the comparative studies between our proposed tool and other available existing tools. Finally, in Section 4, we conclude the paper and state the future work for the proposed tool.

2. Methodology

The research methodologies for enriching webpages with semantic information are divided into two studies. The first study investigates a method for generating a new webpage with microdata; and, the second study investigates a method for enhancing an existing webpage with microdata.

In the first study, we generate a new webpage with microdata. We begin by modeling patent information and representing the patent metadata in the RDF formalism. This is followed by information extraction from patent knowledge base, metadata generation and webpage annotation. Finally, a new HTML webpage is composed with microdata to produce a semantically annotated webpage.

In the second study, we aim to enrich an existing webpage with microdata. The processes involved include: entity recognition, metadata generation and webpage annotation. The entity recognition process uses a Text Annotation Engine called T-ANNE (Xian et al., 2012) that utilizes the Named Entity Recognition technique developed by Nadeau (2007), to identify any semantic entities in the webpage that match the entities from the patent knowledge base. In general, T-ANNE is a knowledge-based solution for entity recognition that is able to identify any semantic entities from a given knowledge base. It matchs semantic entities that occur in the text with the resources stored in knowledge base and returns the references (URI) of the identified resources (Xian et al., 2012).

2.1 Generating new Webpages with Semantic Information

2.1.1. Patent Information Knowledge Modeling

The modeling of patent knowledge begins with the conceptualization of the patent information concepts, properties and individuals and representing them using semantic technology—i.e., RDF (Manolo et al., 2004), RDFS (Brickley et al., 2004) and OWL[1] (Smith et al., 2004). The Spiral Modeling technique (Mohamed et al., 2010) is used to model the ontology and Top-Braid Composer (TopQuadrant, 2007) is used to engineer the ontology. The Schema.org terms are represented using RDFS vocabulary in schema.rdfs.org (Hausenblas and Cyganiak, 2011). We use schema.rdf.org in our ontology to represents the metadata about patent information. Figures 1 and 2 depict some of the resources in the ontology.

The patent knowledge base stores three main groups of information. They are patent document, assignee, and inventor. Patent document has the following set of properties: *title*, *abstract*, *claims*, *description*, *patent number*, *application number*, *filing date*, *publication date*, *copyright year*, *creator (inventor)*, *copyright holder (assignee)*, *examiner*, *attorney*, *reference cited*, and *patent classification*. Assignee, has a set of properties: *company name*, *company address*, *patent owned* and *co-filing activity*, while the properties such as *inventor name*, *inventor address*, *coauthorship*, and *patent invented* is associated with the inventor group. Although we use a lot of schema.org terms in the patent knowledge base, there are still a few specific terms that do not exist in schema.org, but are necessary for our modeling purpose. For instance, we define a *hasClaim* property to describe patent claims and a *hasReferenceCited* property to describe patent citation.

[1] We use OWL version 2 - http://www.w3.org/2002/07/owl

https://doi.org/10.23106/dcmi.952136024

Figure 3 depicts an example of patent information and how the data is structured and represented in our Patent knowledge base[2].

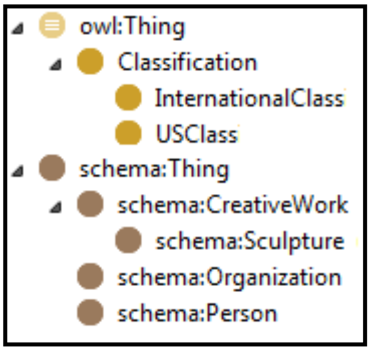


FIG.1. Overview of the Classes in Patent Ontology

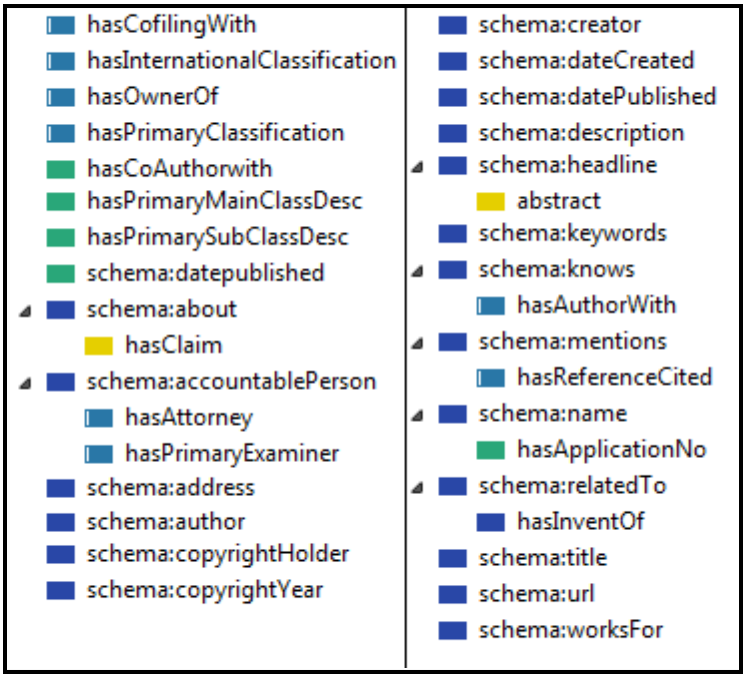


FIG.2. Overview of the Properties in Patent Ontology

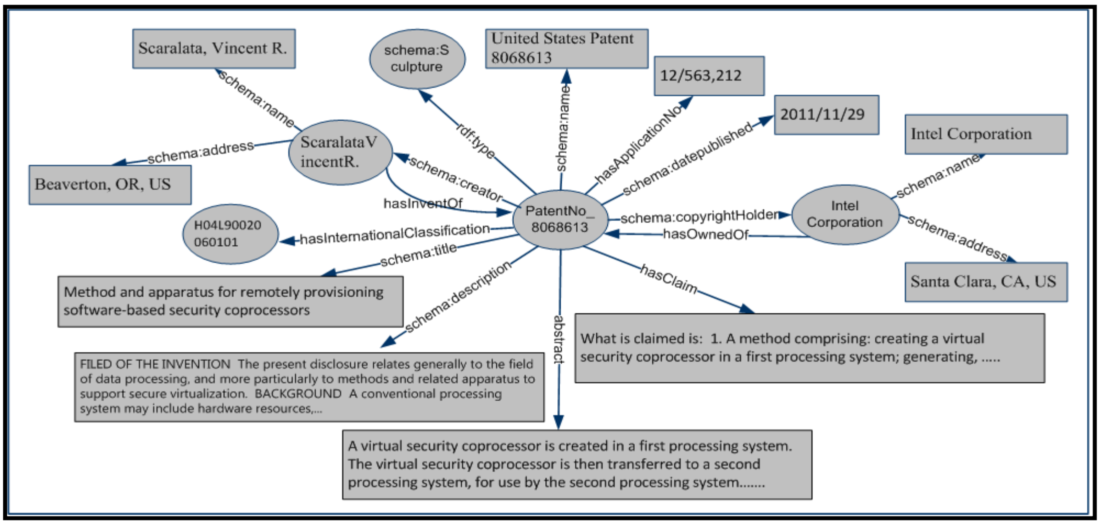


FIG.3. Metadata of Patent Number 8068613

2.1.2. Generating new webpage with Microdata

The ScheMicCr tool is made up of two main components. The first component, which is used for generating a new webpage with microdata, has two processes: information extraction and semantic annotation process. Figure 4 illustrates these processes. The first process extracts the candidate items and their attributes from Patent Knowledge Base. For example, the system

[2] Only partial information is depicted in the diagram for simplified illustration. Other information that is not depicted in the diagram are patent number, examiner, attorney , filing date and copyright year, reference cited, co-filing, coauthorship, and cooperationship.

https://doi.org/10.23106/dcmi.952136024

extracts each patent document and the associated attributes such as title, description, creator, copyright holder, claims and publication date. Next, in the second process, these data are passed to the Microdata Generator, which generates microdata tags. Then, the Webpage Annotation Processor renders predefined html with the microdata tags to produce a semantically annotated webpage.

We execute the process using the example shown in Figure 3. Figure 6 depicts the result: a partial html source annotated with microdata tags. The information in a web browser will not be affected, as illustrated in Figure 5, although the html source has been enriched with microdata.

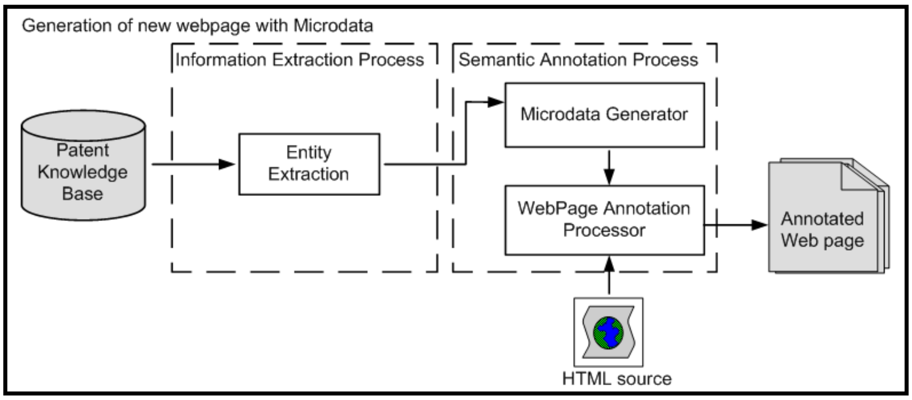


FIG.4. Overall process of generating new webpage with microdata

Patent No 8068613

Title	Method and apparatus for remotely provisioning software-based security coprocessors
Abstract	A virtual security coprocessor is created in a first processing system. The virtual security coprocessor is then transferred to a second processing system, for use by the second processing system.......
Inventor	Vincent R Scaralata
Assignee	Intel Corporation
Publication Date	29/11/2011
Application Number	12/563,212

FIG.5. Information for Patent No 8068613 displayed in a web browser

https://doi.org/10.23106/dcmi.952136024

```html
<html>
<itemscope itemtype="http://schema.org/Sculpture">
    <head><span itemprop="name"> United States Patent 8068613 </span></head>
    <body><table border="1">
 <tr> <td>Title</td>
    <td><span itemprop="title">Method and apparatus for remotely provisioning software-based security coprocessors</span></td>
     </tr>
  <tr> <td>Abstract</td>
    <td><span itemprop="http://mimos.my/tia/abstract">A virtual security coprocessor is created in a first processing system.
    The virtual security coprocessor is then transferred to a second processing system, for use by the second processing
    system…….</span></td> </tr>
 <tr> <td>Inventor</td>
 <span itemprop="creator"> <itemscope itemtype="http://schema.org/Person">
 <td><span itemprop="name">Scaralata, Vincent R.</span></td>
 </span> </tr>
  <tr> <td>Assignee</td>
 <span itemprop="copyrightHolder"> <itemscope itemtype="http://schema.org/Organization">
 <td><span itemprop="name">Intel Corporation</span></td> </span> </tr>
  <tr> <td>Publication Date</td>
 <td><span itemprop="datepublished">29/11/2011</span></td> </tr>
  <tr> <td>Application Number</td>
 <td><span itemprop="http://mimos.my/tia/hasApplicationNo">12/563,212</span></td> </tr>
 </table>
    </body>
</html>
```

FIG.6. A generated HTML source with microdata

2.2. Enriching existing webpage with Microdata

To enrich an existing webpage with microdata, we developed another component that performs semantic annotation, as illustrated in Figure 7. The process starts with T-ANNE. It identifies semantic entities that occur in the existing webpage (text); associates these entities to the resources stored in patent knowledge base and returns the references (URI) of the identified entities. Next, these references (URI) are passed to the Microdata Generator to generate the microdata tags. Finally, Webpage Annotation Processor replaces the identified entity in the existing webpage with the microdata tags to produce a semantically annotated webpage.

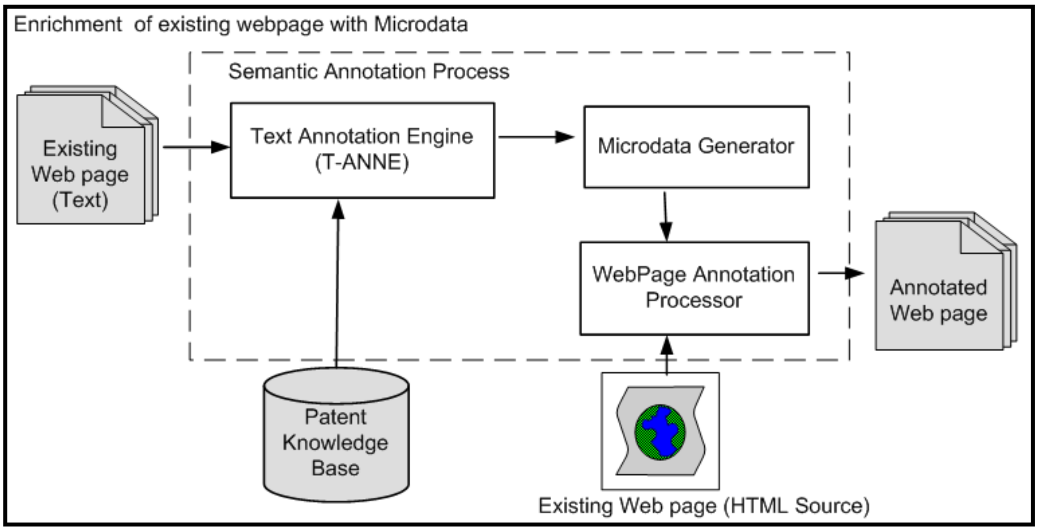


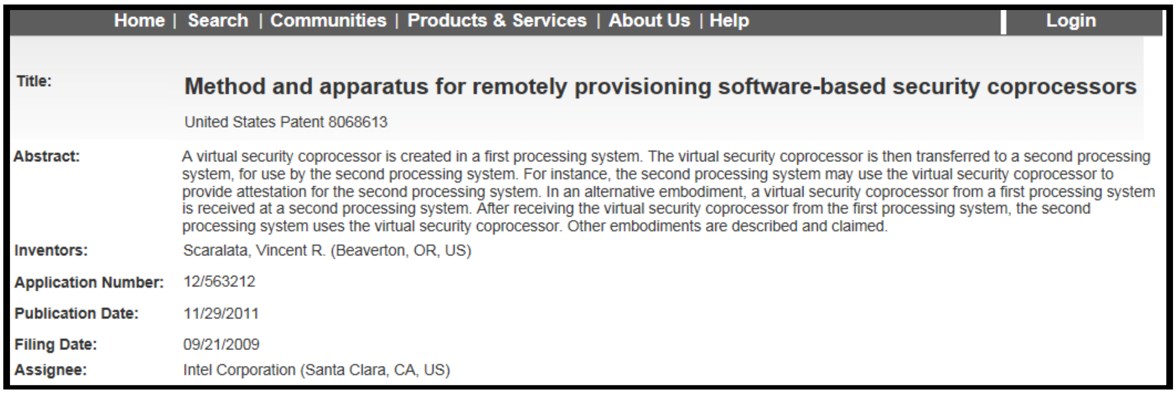
FIG.7. Overall process of enriching existing webpage with Microdata

We are using two types of webpage for our experimental work. The first type is a structure-text webpage from the freepatentonline[3] website. Figure 8 depicts a snapshot of a structured webpage for Patent Number 8068613 displayed in a web browser. For the purpose of a simplified illustration, we only show the information about the title, abstract, inventor, application number,

[3] www.freepatentonline.com

https://doi.org/10.23106/dcmi.952136024

publication date, filing date, and assignee. The snippet of partial original HTML source is depicted in Figure 9 and the result of annotated webpage is depicted in Figure 10.



Home | Search | Communities | Products & Services | About Us | Help | Login

Title:	**Method and apparatus for remotely provisioning software-based security coprocessors** United States Patent 8068613
Abstract:	A virtual security coprocessor is created in a first processing system. The virtual security coprocessor is then transferred to a second processing system, for use by the second processing system. For instance, the second processing system may use the virtual security coprocessor to provide attestation for the second processing system. In an alternative embodiment, a virtual security coprocessor from a first processing system is received at a second processing system. After receiving the virtual security coprocessor from the first processing system, the second processing system uses the virtual security coprocessor. Other embodiments are described and claimed.
Inventors:	Scaralata, Vincent R. (Beaverton, OR, US)
Application Number:	12/563212
Publication Date:	11/29/2011
Filing Date:	09/21/2009
Assignee:	Intel Corporation (Santa Clara, CA, US)

FIG.8. Snapshot of structure-text webpage for Patent No 8068613 from www.freepatentonline.com

```
<!--    TITLE  -->
<div class="disp_doc2"> <div class="disp_elm_title">Title:</div>
    <div class="disp_elm_text"> <font size="+1"><b> Method and apparatus for remotely
     provisioning software-based security coprocessors </b></font>  </div></div>
<!--    Document Type   -->
<div class="disp_doc2"> <div class="disp_elm_text" style="clear: none;"> United States Patent 8068613 </div></div>
<!--    Abstract  -->
<div class="disp_doc2"> <div class="disp_elm_title">Abstract:</div>
    <div class="disp_elm_text">
A virtual security coprocessor is created in a first processing system. The virtual security coprocessor
is then transferred to a second processing system, for use by the second processing system. For instance,
the second processing system may use the virtual security coprocessor to provide attestation for the
second processing system. In an alternative embodiment, a virtual security coprocessor from a first processing system
is received at a second processing system. After receiving the virtual security coprocessor from the first
processing system, the second processing system uses the virtual security
coprocessor. Other embodiments are described and claimed.    </div></div>
<!--    Inventors  -->
<div class="disp_doc2"> <div class="disp_elm_title">Inventors:</div>
    <div class="disp_elm_text"> Scaralata, Vincent R. (Beaverton, OR, US) <br /></div></div>
<!--    Application Number  -->
<div class="disp_doc2"> <div class="disp_elm_title">Application Number:</div>
    <div class="disp_elm_text"> 12/563212 </div></div>
<!--    Publication Date  -->
<div class="disp_doc2"> <div class="disp_elm_title">Publication Date:</div>
    <div class="disp_elm_text"> 11/29/2011 </div></div>
<!--    Filing Date  -->
<div class="disp_doc2"> <div class="disp_elm_title">Filing Date:</div>
    <div class="disp_elm_text"> 09/21/2009 </div></div>
<!--    Assignee  -->
<div class="disp_doc2"> <div class="disp_elm_title">Assignee:</div>
    <div class="disp_elm_text"> Intel Corporation (Santa Clara, CA, US)<br /></div></div>
```

FIG.9. Snippet of original structure-text HTML Source (without microdata)

https://doi.org/10.23106/dcmi.952136024

```
<itemscope itemtype="http://schema.org/Sculpture">
<!--    TITLE  -->
<div class="disp_doc2"> <div class="disp_elm_title">Title:</div>
    <div class="disp_elm_text"> <font size="+1"><b> <span itemprop="title">Method and apparatus for remotely provisioning
    software-based security coprocessors </span>  </b></font> </div></div>
<!--    Document Type   -->
<div class="disp_doc2"> <div class="disp_elm_text" style="clear: none;"> <span itemprop="name">United States Patent 8068613
</span> </div></div>
<!--    Abstract  -->
<div class="disp_doc2"> <div class="disp_elm_title">Abstract:</div>
    <div class="disp_elm_text"><span itemprop="http://mimos.my/tia/abstract">
A virtual security coprocessor is created in a first processing system. The virtual security coprocessor
is then transferred to a second processing system, for use by the second processing system. For instance,
the second processing system may use the virtual security coprocessor to provide attestation for the
second processing system. In an alternative embodiment, a virtual security coprocessor from a first processing system
is received at a second processing system. After receiving the virtual security coprocessor from the first
processing system, the second processing system uses the virtual security
oprocessor. Other embodiments are described and claimed.  </span></div></div>
<!--    Inventors  -->
<div class="disp_doc2"> <div class="disp_elm_title">Inventors:</div>
    <div class="disp_elm_text"> <span itemprop="creator"> <itemscope itemtype="http://schema.org/Person">
    <span itemprop="name"> Scaralata, Vincent R. </span> (Beaverton, OR, US) </span> <br /></div></div>
<!--    Application Number  -->
<div class="disp_doc2"> <div class="disp_elm_title">Application Number:</div>
    <div class="disp_elm_text"> <span itemprop="http://mimos.my/tia/hasApplicationNo"> 12/563212 </span></div></div>
<!--    Publication Date  -->
<div class="disp_doc2"> <div class="disp_elm_title">Publication Date:</div>
    <div class="disp_elm_text"> <span itemprop="datepublished"> 11/29/2011 </span> </div></div>
<!--    Filing Date  -->
<div class="disp_doc2"> <div class="disp_elm_title">Filing Date:</div>
    <div class="disp_elm_text"> <span itemprop="datecreated"> 09/21/2009 </span></div></div>
<!--    Assignee  -->
<div class="disp_doc2"> <div class="disp_elm_title">Assignee:</div>
    <div class="disp_elm_text"> <span itemprop="copyrightHolder"> <itemscope itemtype="http://schema.org/Organization">
    <span itemprop="name"> Intel Corporation </span>(Santa Clara, CA, US) </span><br /></div></div>
```

FIG.10. Snippet of enriched structure-text HTML Source (with microdata)

The second type of experimental webpage used is the free-text (natural language) webpage. Figure 11 depicts the snippet of original HTML source while the result of the annotated webpage is shown in Figure 12.

```
<!doctype html>
<title></title>
<body>
Intel Corporation has granted a United States Patent No 8068613 on 11/29/2011 which was been
submitted on 09/21/2009 . The patent entitled "Method and apparatus for remotely provisioning
software-based security coprocessors" has application number 12/563212 and was invented by Scaralata
Vincent R .
</body>
</html>
```

FIG.11. Snippet of original free-text HTML Source (without microdata)

```
<!doctype html>
<title></title>
<body>
<div itemscope itemtype="http://schema.org/Sculpture">
    <span itemprop="copyrightHolder">
    <span itemscope itemtype="http://schema.org/Organization">
    <span itemprop="name">Intel Corporation </span> </span> </span> has granted a <span itemprop=
    "name"> United States Patent No 8068613 </span> on <span itemprop="datepublished"> 11/29/2011
    </span> which was been filed on <span itemprop="datecreated"> 09/21/2009.</span>
    The patent entitled "<span itemprop="title">Method and apparatus for remotely provisioning
    software-based security coprocessors</span>" has application number <span itemprop=
    "http://mimos.my/tia/hasApplicationNo"> 12/563212 </span> and was invented by <span itemprop=
    "creator"> <span itemscope itemtype="http://schema.org/Person"> <span itemprop="name">Scaralata
    Vincent R </span> </span> </span>.
</div>
</body>
</html>
```

FIG.12. Snippet of enriched free-text HTML Source (with microdata)

https://doi.org/10.23106/dcmi.952136024

2.3. Microdata Validation

There exist a few tools that can perform microdata validation such as Validatornu[4] and RDFa 1.1 Validator[5]. For our experiment, we utilize Validatornu to determine if valid syntax was used in the generated microdata. Figure 13 shows the validation result of generated microdata depicted from Figure 12 (free-text webpage)[6]. It shows that the microdata syntax is valid.

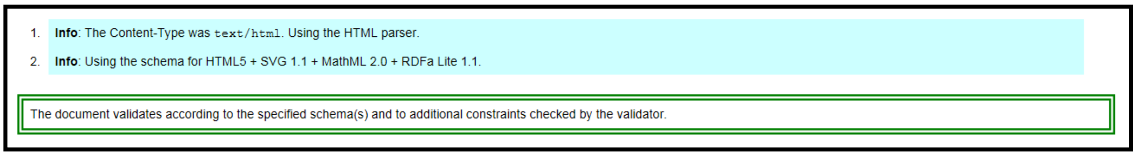


FIG.13. Microdata validation result from Validatornu

Although we do not focus on how the semantic information can be extracted from the annotated HTML webpage, there are a lot of publically available software libraries that can be used to extract the semantic details. These libraries and plugin serve as Microdata parser and are able to process the resulting RDF. Some of these libraries are listed in the Table 1.

TABLE 1: Microdata parser library

Library Name	Script Language	Reference URL
AnythingToTriples (Any23)	Java	http://developers.any23.org/
MicrodataJS	Javascript	http://gitorious.org/microdatajs
MicrodataPHP	PHP	https://github.com/linclark/MicrodataPHP
RDF-MicrodataGem	Ruby	https://github.com/gkellogg/rdf-microdata
Mida	Ruby	http://lawrencewoodman.github.com/mida/
rdflib	Python	https://github.com/edsu/rdflib-microdata

3. Comparison Studies

The comparative studies outlined in this section have been carried out from the perspective of generating webpages with semantic information. We conducted two comparative studies between ScheMicCr and five existing tools. The first study is related to the functionality coverage of the tools. The features we explored include ability to generate new webpages with semantically annotated information, ability to annotate existing webpages, to utilise elements from Schema.org in the microdata generation, to perform RDF to Microdata conversion and to perform Named Entity Recognition. Table 2 lists down the results of the comparative study on functionalities of the tools. From the result, all tools have a feature to generate new webpages with semantic information and utilize schema.org in microdata generation. However, none of the tools, except ScheMicCr, are able to annotate existing webpages and utilise entity recognition technique to identify the entities that will be used for annotation purpose.

[4] http://validator.nu/
[5] http://www.w3.org/2012/pyRdfa/Validator.html
[6] All the generated microdata has been tested and validated. We only show the result of free-text microdata in the diagram for illustration purpose.

https://doi.org/10.23106/dcmi.952136024

TABLE 2: Feature Comparison

Feature/ Tool	New webpage generation	Existing webpage enrichment	SchemaOrg-Microdata generation	RDF to Microdata conversion	Utilize Named Entity Recognition
ScheMicCr	Yes	Yes	Yes	Yes	Yes
microDATA Generator[7]	Yes	No	Yes	No	No
RDF2Microdata Converter[8]	Yes	No	Yes	Yes	No
SchemaCreator[9]	Yes	No	Yes	No	No
SchemFied[10]	Yes	No	Yes	No	No
HTML5 Microdata Template[11]	Yes	No	Yes	No	No

The second study focuses on the qualitative features among the six tools. The qualitative areas that we look into are flexibility, expressiveness, performance, accuracy, and advanced features as shown in Table 3 where the feature metrics are based on our impression scale rating generated based on analyzing 7050 data items in 65 patent documents. For example, the "expressiveness" feature is rated 1 for any tool that has *very poor functionalities* in expressing information, 2 for tools that have *poor functionalities* for the same, 3 for tools with *limited functionalities*, 4 for tools with *adequate functionalities*, while a rating of 5 is given to any tools with *extensive functionalities* to express information.

Based on the qualitative Feature Metrics given in Table 3, a comparative study is conducted on all the tools. Results of this study are listed in Table 4. As shown in the spider chart in Figure 14, ScheMicCr performed reasonably well comparing with its counterparts in this study.

TABLE 3: Qualitative Feature Metrics

Qualitative Feature Metrics	1	2	3	4	5
Flexibility (Able to process any type of data)	n/a (not able to measure)	n/a (not able to measure)	Not Flexible and rigid (fixed type of data)	Flexible and accept any type of data	Very flexible
Expressiveness (Express information)	Very poor functionality to express information	Poor functionality to express information	Limited functionality to express information	Adequate functionality to express all information	Wide range functionality to express all information and can be extendable (inference)
Performance (Handling large dataset)	n/a (not able to measure)	No performance issue to handle very small data sets (<200 data)	No performance issue to handle small data sets(200-499 data)	No performance issue to handle medium datasets (500-999 data)	No performance issue to handle large datasets (>1000 data)
Accuracy (Correctness of generated microdata)	<20% correct	<50% correct	<70% correct	<90% correct	100% correct
Advanced Features(Named entity recognition, RDF to Microdata conversion and existing webpage enrichment feature)	n/a (not able to measure)	No advanced features	At least has one advanced features	At least has two advanced features	At least has three advanced features or more

[7] microDATA Generator - http://www.microdatagenerator.com/
[8] RDF2Microdata Converter - http://www.ebusiness-unibw.org/tools/rdf2microdata/
[9] SchemaCreator - http://schema-creator.org
[10] SchemaFied - http://schemafied.com/
[11] HTML5 Microdata Template - http://microdata.freebaseapps.com/

https://doi.org/10.23106/dcmi.952136024

TABLE 4: Qualitative Feature Comparison

Qualitative Feature	**ScheMicCr**	microDATA Generator	RDF2Microdata Converter	Schema Creator	SchemFied	HTML5 Microdata Template
Flexibility	4	3	4	3	3	3
Expressiveness	4	2	4	2	3	1
Performance	5	1	2	1	1	1
Accuracy	4	4	4	4	4	4
Advanced Feature	5	2	3	2	2	2
(Rated 1 to 5) where 1 is Lowest and 5 is Highest						

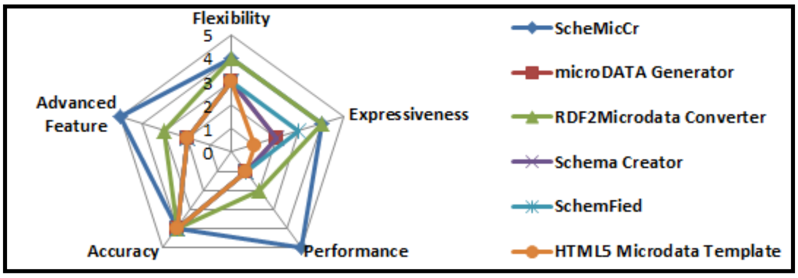


FIG.14. Spider Chart depicting the Qualitative Feature Comparison

4. Conclusion and Future Works

In this paper, we first described how we modeled and engineered the patent knowledge based using the Schema.org terms. We then demonstrated and explained two case studies along with the approach to automatically generation of microdata with Schema.org annotation to enrich html webpages with semantic information. From the qualitative feature comparison, we find that, ScheMicCr significanty leads the other existing tools in terms of its ability to handle mass data (performance) and its advanced features including named entity recognition and existing webpage enrichment feature. For enriching existing webpages, we found our tool to perform adequately as long as the information it is seeking is in the knowledge base. In order words, our tool is totally depended on the knowledge base. Efforts in enriching the knowledge base specific to a particular domain are necessary to ensure high quality annotation.

As for the future work, we will be focusing on benchmarking ScheMicCr by performing deep analysis with large and complex datasets. We will also move from qualitative benchmarking to quantitative benchmarking. We plan to focus on evaluation metrics to improve the system performance by reducing erroneous detection of semantic entities, and incorporating linguistic processing techniques to detect relationships between these entities.

Acknowledgements

We take this opportunity to thank MIMOS Berhad for supporting this research.

References

Brickley, D., and Guha, R.V. (2004). RDF Vocabulary Description Language 1.0: RDF Schema, February 2004. URL: http://www.w3.org/TR/rdf-schema/ (last visited March 2011).

Dobrev, P., and Strupchanska, A. (2005). Conceptual Graphs and Annotated Semantic Web Pages, In Frithjof Dau, Marie-Laure Mugnier, Gerd Stumme (Eds.), Common Semantics for Sharing Knowledge: Contributions to ICCS 2005 13th International Conference on Conceptual Structures, ICCS 2005, Kassel, Germany, ISBN 3-89958-138-5, pp. 54-65. July 1, 2005.

https://doi.org/10.23106/dcmi.952136024

Kudelka, M., Snasel, V., El-Qawasmeh, E., Lehecka, O., and Tesarik, J. (2006) Domain Patterns and Semantic Annotation of Web Pages, Proceedings of the 1st International Conference on Digital Information Management, 6-6 Dec. 2006, pp. 504-510.

Kudelka, M., Snasel, V., Lehecka, O., El-Qawasmeh, and Pokorny, J. (2009). Semantic Annotation of Web Pages Using Web Patterns, Advanced Internet Based Systems and Applications (2009) Volume: 4879, Publisher: Springer Berlin Heidelberg, Pages: 280-291.

Manola, F., and Miller, E. (2004). RDF Primer, February 2004. URL: http://www.w3.org/TR/2004/REC-rdf-primer-20040210/ (last visited March 2011).

Mohamed, S., Mustaffa, S. and Lukose, D. (2010). Using the Spiral Process Model to Develop a Medical Knowledge Base, In Proceedings of the 2nd Semantic Technology and Knowledge Engineering Conference (STAKE 2010), 28-30th July 2010, Kuching, Sarawak, Malaysia, pp. 101-114.

Nadeau, D., and Sekine, S. (2007). A survey of named entity recognition and classification, Journal of Linguisticae Investigationes 30:1; 2007.

Smith, M.K., Welty, C., and McGuinness, D.L. (2004). OWL Web Ontology Language Guide, February 2004. URL: http://www.w3.org/TR/owl-guide/ (last visited March 2011).

Studholme, O. (2011). Using schema.org vocabularies.URL:http://html5doctor.com/microdata/#schema-org-vocab (last visited August 2011)

TopQuadrant (2007). TopBraid Composer: Getting Started Guide (Version 2.0), July 21, 2007.

Xian, B.C.M., Zahari, F., and Lukose, D. (2012).Benchmarking T-ANNE for Named Entity Recognition. In Proceedings of 12th International Conference on Knowledge Management and Knowledge Technologies (i-KNOW).

Hausenblas, M., and Cyganiak, R. (2011). Schema.rdfs.org. URL: http://schema.rdfs.org (last visited December 2011)

https://doi.org/10.23106/dcmi.952136024