International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 4, No.4 (Oct-2015)

http://dx.doi.org/10.12785/ijcds/040406

SAAT: A Manual Annotation Tool for the Arabic Content Authoring

Ahmed N. El-ghobashy[1], Gamal M. Attiya[2], and Hamdy M. Kelash[3]

1, 2, 3 Computer Science & Engineering Department, Faculty of Electronic Engineering, Menoufia University, Egypt.

Received 12 May 2015, Revised 15 July 2015, Accepted 31 August 2015, Published 1 October 2015

Abstract: The Semantic Web (SW) approach actually founded by the actual mass of metadata availability and the representation of data with a proper knowledge. In order to achieve this approach it is necessary to generate metadata that is specific, easy to understand, structured and well defined. Up to now semantic annotation (SA) of a Web document is the effective way to make the SW vision a reality. This paper introduces proof of concept and two case studies of a manual Arabic annotation tool called SAAT that is used for embedding rich metadata within Arabic Web documents to markup the Web pages toward the enrichment process of the Arabic content. This tool is created using the JavaScript and PHP programming language to add Resource Description Framework in Attributes or RDFa metadata for Web pages. The RDFa will make the Arabic Web pages content more structured and machine processable. By using the SAAT tool, we hope to contribute toward the vision of the SW and open the field for Arabic Semantic Web research.

Keywords: Semantic Annotation, Manual annotation tool, RDFa, Structured data, Arabic Language, Semantic content authoring.

1. Introduction

The Semantic Web is an approach that provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries, and facilitate communication by making the web suitable for computers [1]. Tim Berners-Lee created the term for a web of data that can be processed by machines [2].

The Semantic Web envisions a network of semantically enriched content containing links to unambiguous, formal semantics. Even though there is an enormous amount of information, documents and pages accessible online, semantic structured data is not sufficiently supported by search engines (they typically match words syntactically). So far, most of the semantic content available has been created either automatically by wrapping existing data repository or by using annotation facilities for existing content. However, Semantic Web success depends on the accomplishment of a great number of users creating and developing semantic content. This achievement requires tools that reduce the complication of semantic technologies.

One approach for adding a semantic layer to existing web documents is by performing semantic annotation. SA is “the process of labeling Web Pages with the semantics of their contents” [3]. Several Semantic Web annotation tools have been developed recently. These tools have proved their success in multiple domains and different languages, such as English, French and Chinese, to name a few. To transfer this success to the Arabic language, a set of Semantic Web tools needs to be created to deal with the Arabic language contents. In this paper, we present SAAT a tool that provides manual semantic annotation to the Arabic web pages.

The paper is organized as follows: section 2 presents an introduction to Semantic annotation, their categories (manual, semi-automatic and automatic annotation) and a brief explanation of the semantic annotation models and domains. Section 3 presents some existing relate work and their classification, section 4 provides the system architecture of our tool, section 5 highlights two case studies using the tool, section 6 presents the tool evaluation and results, and finally, section 7 gives a

E-mail: ahmed.elghobashy@gmail.com, gamal.attiya@yahoo.com, hmk3947@yahoo.com

conclusion with our future vision for the tool and next improvement.

2. SEMATIC ANNOTATIONS

Nearly few years ago, Tim Berners-Lee the inventor of the WWW [4, 5], speaks about his vision for the Semantic Web and the need for it, since the success of the current web will leads to a new challenge as a massive amount of data is only human understandable; and machine support is partial or absent. Berners-Lee suggests mechanisms to describe data in Semantic Web terms, which will facilitate applications to exploit data in more ways and support the user in his task, one of these ways is the use of semantic annotations.

In general, annotation manner is a remark made though our reading any text. This process may be as simple as underlining or emphasizing passages. Creating these remarks or comments, for limited sentences long, generates a summary for Web content and articulates giving the meaning of each basis. For the most part, an annotation is additional info in a text that categorizes or classifies the meanings of a part of that document.

Therefore, it assigns intellect labels, names, tags, comments, clarifications, etc., to a document or to a particular part in a text. This method helps to companion the vagueness of the natural language (NL) when expressing ideas and their computational demonstration in a proper linguistic, by expressive the computer how data items are related, connected and how these relations can be calculated automatically [6, 7]. For semantic annotation tool, annotations generate specific metadata and usage schema, enabling new information access methods and extending the existing ones. Generating semantic remarks or annotations for a Web resource can be done by using one of the subsequent classifications:

A. *Manual Annotation*

Is considered one of the basic methods for creating semantic annotation, manual tools let the users to add annotations to Web pages or other resources by hand. With the modification of current syntactic resources into added information structures that demonstrate related underlying information structures by adding data to some level of document (word, phrase or paragraph) which constitutes metadata [8]. The process of manual annotation is expensive and time-consuming process. Furthermore, MA is more easily feasible today, by means of authoring tools such as Semantic Word.

As an example of manual annotation, Semantator [9] lets a given user to choose two instances and add them to the relationship applicant list. After that, he can choose any object property from the loaded ontology and decide the subject of this new relationship.

B. *Semi-automatic Annotation*

Semi-automatic tools depends on human involvement at some point for the annotation creation process. The tools diverge in their construction, information extraction and methods, initial ontology, quantity of the manual work necessary to accomplish and create the annotation. As an example for semi-automatic annotation tool is NCBO annotator [10] and cTAKES [11].

C. *Automatic Annotation*

Finally, automatic annotation tools follow two types of schemes that learn how to annotate, these schemes are supervised systems and unsupervised systems. The disadvantage of supervised technique is that choosing satisfactory good patterns is a non-trivial and error-prone task. The other system involves a diversity of strategies to learn how to annotate without any user administration, but their precision is unsatisfactory [12].

Semantic Annotation also can be classified into four models: tags, attributes, relations and ontologies. Tags are placed at the bottom level, correspond to the easiest form of annotation from the user point of view, while ontologies are at the top level, and represent the hardest form of annotation from the user point of view.

- *Tags:* A tag annotation element is a keyword (word or sequence of characters without spaces) or a term assigned to a resource that, implicitly, describes a particular property of a resource.
- *Attributes:* An attribute annotation element is a pair of two elements: the name of the attribute and the value of the attribute. The name of the attribute defines the property of the annotated resource (e.g., “Country”, “City”) and the attribute value specifies the corresponding value (e.g., “Egypt”, “Cairo”).
- *Relations:* a relation annotation element is a pair of two components: the relation name and the related resource. The annotated resource is related with relation by the relation name.
- *Ontologies:* The ontology model describes the metadata that align a resource or a part of it with some of its properties and characteristics description according to a formal conceptual model (ontology).

Table 1 reviews the benefits and drawbacks of eachsemantic annotation classification.

TABLE I. BENEFITS AND DRAWBACKS OF SEMANTIC ANNOTATION METHODS

Method	Benefits	Drawbacks
Manual annotation	Precise method of annotating resources. Sustenance the requirements of different users.	Time consuming method, and often does not consider numerous perceptions of a data source, needful of many ontologies.
Semi-automatic annotation	Suitable speed of annotation with middle precision.	Generated output needs to be revised to make sure it is annotation process is precise.
Automatic annotation	Fast annotation speed method.	Restricted to the usage by specialists while others are suitable for understanding workers. User interface (UI) strategy concerns related with decreasing intrusiveness while get the maximum out of precision.

3. RELATED WORK

Annotation tools that targets metadata annotations can have several formats, this section gives a description and an example for the several formats (e.g. ontological, image and textual) that we can find.

A. Ontology annotation tools

Ontology based document annotation has been the focus of many projects and applications, since the availability of annotated content is one of the key challenges to overcome in order to make the semantic web a reality. Based on the principles of tools that capture the prerequisite of providing clear proper meaning to annotations, the following tools ware selected:

- *GATE* [13, 14]: one of the most broadly recognized systems over the last years. Used for vast organization and text annotation. Comes as a desktop application codes in Java programing language and can be run under nearly any Operating Systems (OS).
- *KIM* [15]: is a platform with a knowledge and information management infrastructure and services for automatic semantic annotation, indexing, and retrieval of unstructured and semi-structured content. Inside the process of annotation, KIM also does ontology population. As a base line, KIM inspects texts and recognizes references to entities (like persons, organizations, locations, dates), then it tries to match the reference with a recognized entity, having a unique Uniform Resource Identifier (URI) and description in the knowledge base [16].
- *Melita* [17]: is a tool developed to define and develop an ontology-based annotation services. It is a semi-automatic annotation tool based on the Amilcare Information Extraction Engine [18].

B. Image annotation tools

In this process, we can assign metadata in the form of captioning or keywords to a digital image by annotation tool, after that we can use an image retrieval system to organize and locate images of interest from a database. The following tools ware selected:

- *M-OntoMat-Annotizer* [19]: allows the semantic annotation of images for multimedia analysis and retrieval. It is an extension of the CREAM (Creating Metadata for the Semantic Web) framework and its reference implementation, OntoMat-Annotizer. The Visual Descriptor Extraction Tool (VDE) developed as a plug-in to OntoMat-Annotizer and is the core component for extending its capabilities and supporting the initialization and linking of RDF(S) domain ontologies with visual descriptors.
- *SpiritTagger* [20]: Is a system of image annotation invented to explore knowledge extraction through mining of millions of global photographs referenced with a geographical coordinate.
- *SML* [21]: produces a natural ordering of semantic labels at annotation time, and eliminates the need to compute a "non-class" model for each of the semantic concepts of interest.

C. Text annotation tools

These tools focus to transform the existing syntactic resources into interlinked knowledge structures that represent relevant underlying information. Our main concern is the authoring tools. Following is a few examples:

- *Amaya* [22]: an interactive Web browser and editor with user-friendly interface built on the Annotea framework, which can mark-up Web documents in extensible Mark-up Language (XML) and HTML. The user can generate annotations with the same tool they use for surfing and for editing text, making Amaya a respectable example of a single point of access environment.
- *OCA* [23]: One Click Annotation is a Web editor that follow the principle of What You See Is What You Get (WYSIWYG). OCA used with Web browsers to allows for annotating words and phrases with references to ontology concepts and for generating relations between annotated phrases by enriching content with Resource Description Framework in Attributes (RDFa) annotations. It has an intuitive user interface that

aims to hides the difficulty of producing semantic data. The greatest thing about OCA is they consider non-experts with little or no familiarity with semantic technologies as the main target group, which is a novel way needed to the success of the SW annotation since it depends on achievement an enormous amount of users generating, creating and consuming semantic content [24].

- *AraTation* [25]: is an Arabic semantic annotation tool for focusing in the field of Arabic News content annotation on the Web. Applied as a desktop application. AraTation tool created using Java and Web Ontology Language (OWL) to output Resource Description Framework (RDF) metadata for Web pages.

4. SAAT

During the last few years, many semantic Web annotation tools have been developed, however a very limited number considered working with Arabic contents. With the use of SAAT, we aim to provide a tool for annotating Arabic content semantically. Our focus to provide a tool which is suitable for non-expert users having slight or no familiarity of semantic technologies, because we consider that the success of the Semantic Web depends on getting a critical mass of users creating and consuming semantic content.

A. SAAT Architecture

In our last work [26], we proposed a framework for Arabic semantic annotation tool with some features that can be used in the development process of the tool, to continue our work we have developed SAAT as a proof of concept.

SAAT is a plugin implemented over a content management system or CMS [27], the main role of the tool is to create, manage, and access semantically enriched content, likewise to CMS allowing people unfamiliar with HTML to manage the content of websites, SAAT allows people unfamiliar with semantic technologies to add semantic annotations to Arabic texts. The client side is responsible for adding, editing and publishing the annotated Arabic content, where the server side is only used in serializing and retrieving the annotated web page contents. The main core of SAAT is the vocabularies from Schema.org [28] that we used the classes and its attributes in annotating the content. Fig 1. Shows a diagram for the tool architecture.

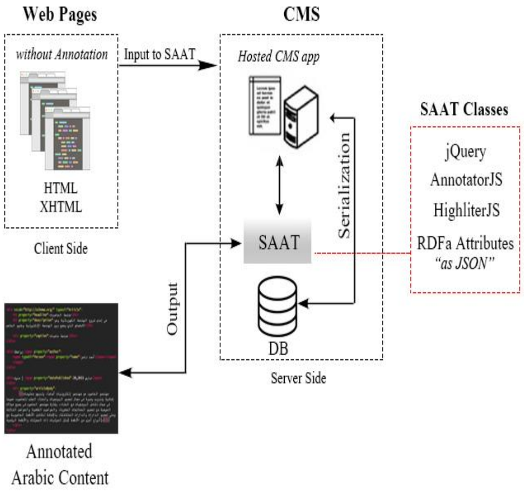
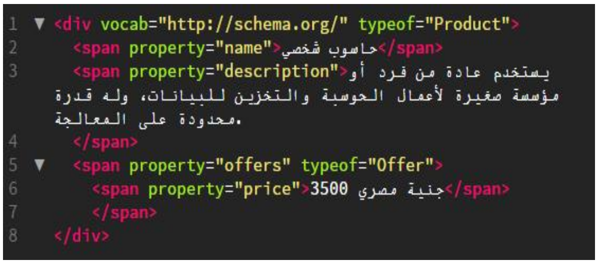


Figure 1. SAAT architcture

The procedure of annotating a content of a web page is as follows:

- Loading the web page which requires annotation into the CMS editor area
- Selecting the desired text for annotation and click the right mouse button
- A context-menu will appear with only two options the class name and the attributes for .the selected class, e.g. a class name "Product / منتج "
- Now we can add the class attributes to the highlighted area of the selected text which was given a class before, e.g. "name/الاسم", "description / الوصف" and "price / السعر"

Fig 2. Shows a snippet of the generated code stating that there is a resource "Product / منتج" with the name "personal computer /حاسوب شخصي", "description / الوصف" and a "price /السعر".



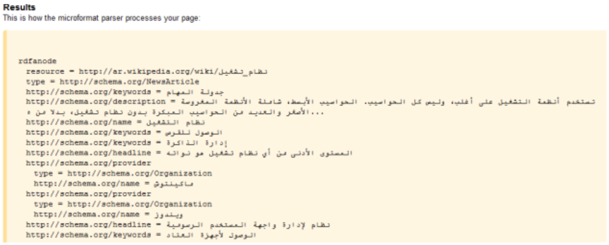
```
<div vocab="http://schema.org/" typeof="Product">
  <span property="name">حاسوب شخصي</span>
  <span property="description">يستخدم عادة من فرد أو مؤسسة صغيرة لأعمال الحوسبة والتخزين للبيانات، وله قدرة محدودة على المعالجة.
  </span>
  <span property="offers" typeof="Offer">
    <span property="price">3500 جنية مصري</span>
    </span>
</div>
```

Figure 2. SAAT generated code snippet

B. SAAT evaluation

This section shows an evaluation to the output of the tool, however many evaluation aspects need to be considered, for now we will concentrate to check the validation, parse of the generated markup and the relations between the resources. The evaluation was performed on 20 different Arabic texts, as an example we will highlight two cases test material texts from the Arabic Wikipedia about "operating systems / أنظمة التشغيل" [29] and "Egypt/مصر" [30].

1) Case: "operating systems /أنظمة التشغيل ": we annotated the content with "News Article" vocabularies and its class attributes, when we validate the content according to Yandex [31] a Webmaster Tool for parsing RDFa content we got a full understand of the annotated contents as displayed in Fig 3.

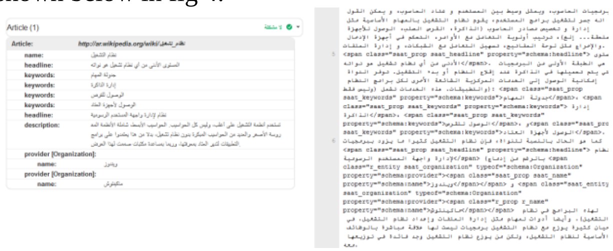


Results
This is how the microformat parser processes your page:

```
rdfanode
  resource = http://ar.wikipedia.org/wiki/نظام_تشغيل
  type = http://schema.org/NewsArticle
  http://schema.org/keywords = جدولة المهام
  http://schema.org/description = تستخدم أنظمة التشغيل على أغلب، وليس كل الحواسيب. الحواسيب الأبسط، شاملة الأنظمة المغروسة
  ء من بدلا ،تشغيل نظام بدون المبكرة الحواسيب من والعديد الأصغر...
  http://schema.org/name = نظام التشغيل
  http://schema.org/keywords = الوصول للقرص
  http://schema.org/keywords = إدارة الذاكرة
  http://schema.org/headline = المستوى الأدنى من أي نظام تشغيل هو نواته
  http://schema.org/provider
    type = http://schema.org/Organization
    http://schema.org/name = ماكينتوش
  http://schema.org/provider
    type = http://schema.org/Organization
    http://schema.org/name = ويندوز
  http://schema.org/headline = نظام لإدارة واجهة المستخدم الرسومية
  http://schema.org/keywords = الوصول لأجهزة العتاد
```

Figure 3. Yandex parsing RDFa results

In addition, we validated according to Google Webmaster Tools [33] for testing the validation of the relation between the resource and its attribute, we passed the entire validation test without any problems mentioned as shown below in fig 4.



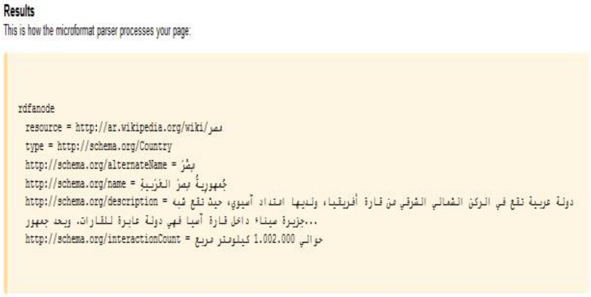
Article (1) لا مشكلة

Article: http://ar.wikipedia.org/wiki/نظام_تشغيل
name: نظام التشغيل
headline: المستوى الأدنى من أي نظام تشغيل هو نواته
keywords: جدولة المهام
keywords: إدارة الذاكرة
keywords: الوصول للقرص
keywords: الوصول لأجهزة العتاد
headline: نظام لإدارة واجهة المستخدم الرسومية
description: تستخدم أنظمة التشغيل على أغلب، وليس كل الحواسيب. الحواسيب الأبسط شاملة الأنظمة المغ روسة الأصغر والعديد من الحواسيب المبكرة بدون نظام تشغيل، بدلا من هذا يعتمدوا على برامج التطبيقات لتدير العتاد بمفردها، وربما بمساعدة مكتبات صممت لهذا الغرض.
provider [Organization]:
name: ويندوز
provider [Organization]:
name: ماكينتوش

```
برمجيات الحاسوب، ويمثل وسيط بين المستخدم و عتاد الحاسوب، و يمكن القول
انه جسر لتشغيل برامج المستخدم، يقوم نظام التشغيل بالمهام الأساسية مثل
إدارة و تخصيص مصادر الحاسوب (الذاكرة، القرص الصلب، الوصول للأجهزة
ملحقة... إلخ)، ترتيب أولوية التعامل مع الأوامر، التحكم في أجهزة الإدخال
والإخراج مثل لوحة المفاتيح، تسهيل التعامل مع الشبكات، و إدارة الملفات.
<span class="saat_prop saat_headline" property="schema:headline"> ستوى
ي الطبقة الأولى من البرمجيات .الأدنى من أي نظام تشغيل هو نواته</span>
تي يتم تحميلها في الذاكرة عند إقلاع النظام أو بدء التشغيل. توفر النواة
إمكانية الوصول إلى الخدمات المركزية الشائعة الأخرى لكل برامج النظام
والتطبيقات. هذه الخدمات تشمل (وليس فقط): <span class="saat_prop
saat_keywords" property="schema:keywords">جدولة المهام</span>، <span
class="saat_prop saat_keywords" property="schema:keywords"> إدارة
الذاكرة</span>، <span class="saat_prop saat_keywords"
property="schema:keywords">الوصول للقرص</span>، و<span class="saat_pro
saat_keywords" property="schema:keywords">الوصول لأجهزة العتاد</span>.
كما هو الحال بالنسبة للنواة، فإن نظام التشغيل كثيرا ما يزوده ببرمجيات
<span class="saat_prop saat_headline" property="schema:headline"> نظام
لإدارة واجهة المستخدم الرسومية</span> (بالرغم من إدماج <span
class="r_entity saat_organization" typeof="schema:Organization"
property="schema:provider"><span class="saat_prop saat_name"
property="schema:name">ويندوز</span></span> و <span class="saat_entity
saat_organization" typeof="schema:Organization"
property="schema:provider"><span class="r_prop r_name"
property="schema:name">ماكينتوش</span></span> لهذه البرامج في نظام
التشغيل)، وأيضا أدوات لمهام مثل إدارة الملفات وإعداد نظام التشغيل. في
حيان كثيرة يوزع مع نظام التشغيل برمجيات ليست لها علاقة مباشرة بالوظائف
الأساسية لنظام التشغيل، ولكن من يوزع نظام التشغيل وجد فائدة في توزيعها
معه.
```

Figure 4. Google Webmaster generated markup and resource relations

2) Case: "Egypt /مصر": we annotated the content with "Country" vocabularies and its class attributes, when we validate the content according to Yandex we got a full understand of the annotated contents as in fig 5.

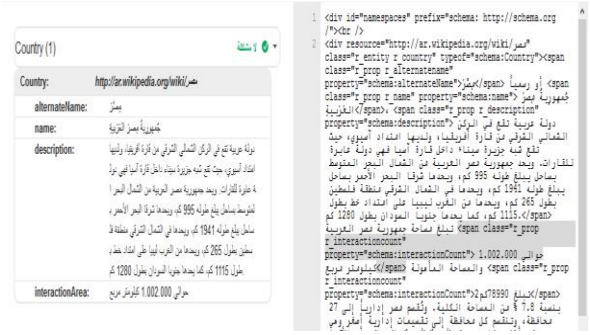


Results
This is how the microformat parser processes your page:

```
rdfanode
  resource = http://ar.wikipedia.org/wiki/مصر
  type = http://schema.org/Country
  http://schema.org/alternateName = مِصْر
  http://schema.org/name = جُمهوريةُ مِصرَ العَرَبيةِ
  http://schema.org/description = دولة عربية تقع في الركن الشمالي الشرقي من قارة أفريقيا، ولديها امتداد آسيوي، حيث تقع شبه
  جمهور ويحد .للقارات عابرة دولة فهي آسيا قارة داخل سيناء جزيرة...
  http://schema.org/interactionCount = حوالي 1.002.000 كيلومتر مربع
```

Figure 5. Yandex parsing RDFa results

Also as the previous case, we validated according to Google Webmaster Tools for testing the validation of the relation between the resource and its attribute, we passed the entire validation test without any problems mentioned fig 6.



Country (1) لا مشكلة

Country: http://ar.wikipedia.org/wiki/مصر
alternateName: مِصْر
name: جُمهوريةُ مِصرَ العَرَبيةِ
description: دولة عربية تقع في الركن الشمالي الشرقي من قارة أفريقيا، ولديها امتداد آسيوي، حيث تقع شبه جزيرة سيناء داخل قارة آسيا فهي دولة عابرة للقارات. ويحد جمهورية مصر العربية من الشمال البحر المتوسط بساحل يبلغ طوله 995 كم، ويحدها شرقا البحر الأحمر بساحل يبلغ طوله 1941 كم، ويحدها في الشمال الشرقي منطقة فلسطين بطول 265 كم، ويحدها من الغرب ليبيا على امتداد خط بطول 1115 كم، كما يحدها جنوبا السودان بطول 1280 كم
interactionArea: حوالي 1.002.000 كيلومتر مربع

```
<div id="namespaces" prefix="schema: http://schema.org
/"><br />
<div resource="http://ar.wikipedia.org/wiki/مصر"
class="r_entity r_country" typeof="schema:Country"><span
class="r_prop r_alternatename"
property="schema:alternateName">مِصْر</span> أو رسمياً <span
class="r_prop r_name" property="schema:name"> جُمهوريةُ مِصرَ
العَرَبيةِ</span>، <span class="r_prop r_description"
property="schema:description"> دولة عربية تقع في الركن
 الشمالي الشرقي من قارة أفريقيا، ولديها امتداد آسيوي، حيث
 تقع شبه جزيرة سيناء داخل قارة آسيا فهي دولة عابرة
للقارات. ويحد جمهورية مصر العربية من الشمال البحر المتوسط
 بساحل يبلغ طوله 995 كم، ويحدها شرقا البحر الأحمر بساحل
 يبلغ طوله 1941 كم، ويحدها في الشمال الشرقي منطقة فلسطين
 بطول 265 كم، ويحدها من الغرب ليبيا على امتداد خط بطول
1115.</span> كم، كما يحدها جنوبا السودان بطول 1280 كم
تبلغ مساحة جمهورية مصر العربية <span class="r_prop
r_interactioncount"
property="schema:interactionCount"> 1.002.000 حوالي
كيلومتر مربع</span> والمساحة المأهولة <span class="r_prop
r_interactioncount"
property="schema:interactionCount">2كم78990 تبلغ</span>
 بنسبة 7.8 % من المساحة الكلية. وتُقسم مصر إدارياً إلى 27
 محافظة، وتنقسم كل محافظة إلى تقسيمات إدارية أصغر وهي
```

Figure 6. Google Webmaster generated markup and resource relations

5. CONCLUSIONS AND FUTURE WORK

This paper presented an architecture, its implementation and use cases of a manual semantic annotation tool targeting the Arabic content toward leveraging Semantic Web technologies to serve the Arabic language, and produce semantically annotated web pages. Even if the tool existing in this paper has achieved its intended goals and proof of concept, many possible additions can improve the tool's performance and productivity. The enhancements include:

- Add more vocabularies and NE with the use of different approach than Scema.org
- Increasing of the included domains used
- Increasing the performance and the response time of the tool
- Adding support to images and multimedia
- Adding support for keyword extraction

ACKNOWLEDGMENT

I wish to express my sincere thanks to Prof. Ahmed Refaat Al sisi for his kind support and help in writing this paper.

REFERENCES

[1] "W3C Semantic Web Activity". World Wide Web Consortium (W3C). November 7, 2011. Retrieved April 25, 2015.

[2] Berners-Lee, Tim; James Hendler; Ora Lassila (May 17, 2001). "The Semantic Web". Scientific American Magazine. Retrieved (March 26, 2011).

[3] Li, D. and Huan, L. The Ontology Relation Extraction for Semantic Web Annotation. Ccgrid. (2008) Eighth IEEE International Symposium on Cluster Computing and the Grid (CCGRID). 534-541.

[4] Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The semantic web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. In Scientific American.

[5] http://en.wikipedia.org/wiki/Tim_Berners-Lee. Retrieved April 25, 2015.

[6] Handschuh, S. Creating ontology-based metadata by annotation for the semantic web. Doctoral dissertation, Karlsruhe, Univ., Diss., (2005).

[7] Cardoso, J. The semantic web vision: Where are we ? IEEE Intelligent Systems, 22(5), 84-88. (2007).

[8] Dingli, A. Annotation for the Semantic Web. In Knowledge Annotation: Making Implicit Knowledge Explicit (pp. 19-24). Springer Berlin Heidelberg. (2011).

[9] Song, D., Chute, C. G., & Tao, C. Semantator: a semi-automatic semantic annotation tool for clinical narratives. In 10th International SemanticWeb Conference (ISWC2011).

[10] Storey, M. A., Chute, C. G., and Musen, M. A. BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Res, 37 (Web Server issue), 170-173. (2009).

[11] Kipper-Schuler, K.C., Chute. clinical text analysis and knowledge extraction system (ctakes): architecture, component evaluation and applications. Journal of the American Medical Informatics Association 17(5), 507-513. (2010).

[12] Cunningham, K. B. H. 3 Semantic Annotations and Retrieval: Manual, Semiautomatic, and Automatic Generation. Handbook of Semantic Web Technologies, March 1. (2011).

[13] URL: http://gate.ac.uk/ [Last accessed April 30, 2015].

[14] Cunningham, H., Maynard, D., & Bontcheva, K. Text processing with gate. Gateway Press CA. (2011).

[15] URL: https://www.w3.org/2001/sw/wiki/KIM_Platform [Last accessed April 30, 2015].

[16] Popov, B. et al. Kim - semantic annotation platform, in ISWC, 834–849. (2003).

[17] Ciravegna, F. et al. Amilcare: adaptive information extraction for document annotation. In SIGIR, 367–368, (2002).

[18] Ciravegna, F. et al. User-system cooperation in document annotation based on information extraction. In EKAW, 122–137. (2002).

[19] Petridis, Kosmas, et al. "M-ontomat-annotizer: Image annotation linking ontologies and multimedia low-level features." Knowledge-Based Intelligent Information and Engineering Systems. Springer Berlin Heidelberg, (2006).

[20] Moxley, E. Kleban, J. Manjunath, B. S. Spirittagger: a geo-aware tag suggestion tool mined from flickr, Proceedings of the 1st ACM international conference on Multimedia information retrieval, October 30-31, Vancouver, British Columbia, Canada. (2008).

[21] Carneiro, G., Chan, A. B. Moreno, P. J. and Vasconcelos, V. 2006. "Supervised Learning of Semantic Classes for Image Annotation and Retrieval," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 29(3), 394-410. (2006).

[22] URL: http://www.w3.org/Amaya/ [Last accessed May 3, 2015].

[23] Heese, Ralf, et al. "One Click Annotation." SFSW. (2010).

[24] Heese, R., Luczak-Rösch, M., Oldakowski. One click annotation. In Proceedings of the Workshop on Collaborative Construction, Management and Linking of Structured Knowledge (CK). Ed. by Tania Tudorache, Gianluca Correndo, Natasha Noy, Harith Alani, and Mark Greaves. CEUR Workshop Proceedings: http://CEUR-WS.org (Vol. 514). (2010).

[25] Saleh, L. M. B., & Al-Khalifa, H. S. AraTation: an Arabic semantic annotation tool. In Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services (pp. 447-451). ACM. (2009).

[26] El-ghobashy, Ahmed N., Gamal M. Attiya, and Hamdy M. Kelash. "A Proposed Framework for Arabic Semantic Annotation Tool." Int. J. Com. Dig. Sys 3.1 (2014): 47-53.

[27] URL: http://en.wikipedia.org/wiki/Content_management_system [Last accessed May 10, 2015].

[28] URL: http://schema.org/ [Last accessed May 10, 2015].

[29] URL: http://ar.wikipedia.org/wiki/نظام_تشغيل [Last accessed May 10, 2015].

[30] URL: http://ar.wikipedia.org/wiki/مصر [Last accessed May 10, 2015].

[31] URL: https://webmaster.yandex.com/microtest.xml [Last accessed May 10, 2015].

[32] URL: https://developers.google.com/structured-data/testing-tool/ Last accessed May 10, 2015].

Ahmed N. El-ghobashy graduated in 2009 and obtained his B. Sc. degree in computer engineering and information technology and he is now pursuing M.Sc. in computer science and information technology. His main interests include Web development, Web design, Semantic Web, Information architecture, CMS, Programming and web-standards.

Gamal M. Attiya graduated in 1993 and obtained his M.Sc. degree in computer science and engineering from the Menoufia University, Egypt, in 1999. He received his PhD degree in computer engineering from the University of Marne-La-Vallée, Paris-France, in 2004. His main research interests include distributed computing, task allocation and scheduling, computer networks and protocols, congestion control, QoS, multimedia networking and Image processing.

Hamdy M. Kelash received the Eng. degree from the institute of Electronic Engineering, Egypt in 1971. M.Sc. degree from Faculty of Engineering Technology, Helwan University, Egypt in 1979 and the PHD degree from institute National Polytechnique (INP), France in 1984. He has been lecturer since 1984 at the Electronic Industry department, Faculty of Electronic Engineering also a lecturer in 1987at the Computer Science and Engineering department and an Assistant Professor in 1993, and the Head of Computer Sciences and Engineering department Faculty of Electronic Engineering, Menoufia University from 2001 to 2007. His main research interests includes optical computing, artificial intelligence, network security, image processing, digital systems and parallel computing.