Enrichment Ontology with Updated user Data for Accurate Semantic Annotation

Haytham Al-Feel[1], Hanaa Ghareib Hendi[2], Heba Elbeh[3]
Faculty of Computers & Information
Fayoum University, Fayoum, Egypt

***Abstract*—Annotation is considered one of the main applications that semantic web applies. The idea beyond annotation focused on adding metadata to existing information which facilitates machines dealing with data that have meanings and can be readable. Semantic annotation is one of the techniques used for the enrichment of web content semantically, which facilitates writing comments and evaluate previously annotated resources that can lead to better search results. Our framework aims to enrich ontology via embedding data directly to ontology in order to have completed and accurate data.**

***Keywords*—Ontology; Semantic Web; Semantic Annotation; RSS News**

I. Introduction

The current web has many problems and limitations where most of the web pages can't be understood by machines, which makes the web flat, boring and also misunderstandings [1]. Semantic Web has many definitions but most of them refer to it as an extension for the current web that links information to each other. It also could link data to anything that could have a Unified Resource Identifier and in this way, the web will be enabled for accurate retrieval [2]. Absolutely, the work is not easy like this but it will include many sophisticated technologies and languages such as the Resource Description Framework (RDF), Web Ontology Language (OWL) and SPARQL as a query language.

Semantic annotation can be defined as adding metadata to available documents that add value and explanation to the web, facilitates the searching process and finding any piece of information especially when these data resources linked to each other [3] [4] [5]. Annotation is a set of explanatory notes that accompanies a text [6].

Ontology considered one of the main players of the semantic web and the backbone for the annotation process that explained as the formal specification for a specific domain explaining its concepts semantically [7]. Ontology triples have main three parts they are: subject which describes a specific resource and identified by a URI, predicate which identifies a property and an object which can be another resource or value as shown in "Fig. 1". Because of syntax and structure which has a lot of details, specifications, and recommendations, regular users do not have the full awareness of ontology to deal with or insert, update and delete data. The task of annotation could enrich content and ontologies in different domains.

Our paper presents a Semantic annotation tool that facilitates the user adding data to available resources according to specific rules and privileges determined by the admin of each website. Admin reviews the user's insertion, whether true or false. If the user insertion is acceptable, in this case, data inserted will be stored in the ontology triple store to enrich ontology. Users do not need to have any background about semantic and ontology to use our tool and framework, just he/she will insert data that will be transformed in the form of subject, predicate, and object.

The remainder of this paper is organized as follows: Section 2, highlights semantic annotation techniques. While Section 3, presents our framework and evaluation. Finally, Section 4 presents our conclusions and further research.

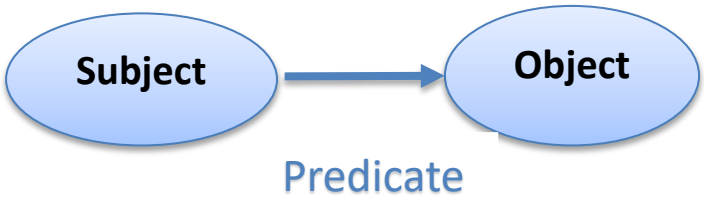

Fig. 1. Ontology Triple Structure.

II. Semantic Annotation

Remarks, comments, notes and any explanations added to a part of web documents are considered annotations to these web documents and if these annotations based on ontologies and semantic web techniques; in this situation, we call them semantically annotated [8]. Semantic annotation is defined as the process of adding data related to ontological concepts to have full metadata description about concepts. The purpose of a semantic annotation is to specify the meaning and properties of an annotated information resource in a generally understandable way [9]. It recognizes the parts of the text and converts into pieces in a data processing mechanism that can be linked to a broader context of data that already exist [10]. In [11] we developed a Semantic Annotation Framework for News Feeds (SANF) which annotates RSS news titles and provides additional information for concepts on the news domain via an ontology contains different categories and concepts in this domain as shown in "Fig. 2". In this paper, we added new features for enhancing the system via adding full authority to the user to add a comment on what retrieved from an ontology which could be later accepted or refused from the administrator. In addition to this, even if a particular user needs a brief annotation of an unannotated object the system does not stand idly; it acts intelligently by looking for an object online in order to perform an annotation process as fully as possible.

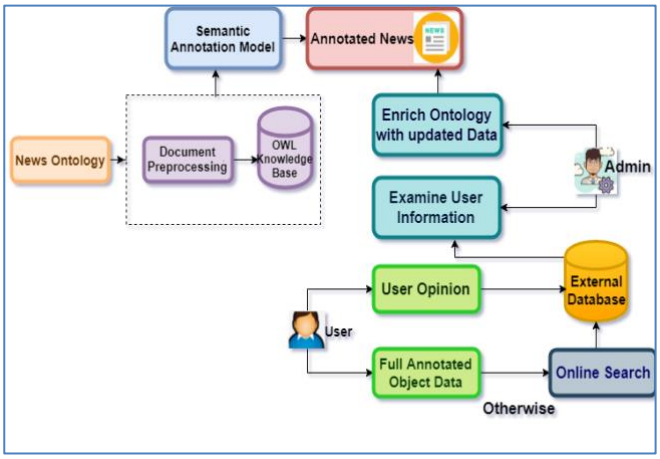

Fig. 2. System Architecture.

III. System Framework

We enhance our proposed system to support the user to have the full ability to insert his/her own metadata into the system repository. After that, the system admin examines and reviews the inserted one in order to enrich ontology with updated data to enrich the data repository.

Each class in the developed ontology has many individuals with its properties and relations which in turn clearly illustrates the components of this class as shown in "Fig. 3".

The system validates URLs entered by users and apply the required process of news titles to finally provide a semantic annotation model to map concepts with its related ontology class. The final stage of the system is an annotated web page of news feeds as shown in "Fig. 4".

If the user selects any of the annotated words, the full annotation would be retrieved then the user can insert his/her reviews about the information retrieved from the ontology as shown in "Fig. 5" where it is stored successfully in an external repository.

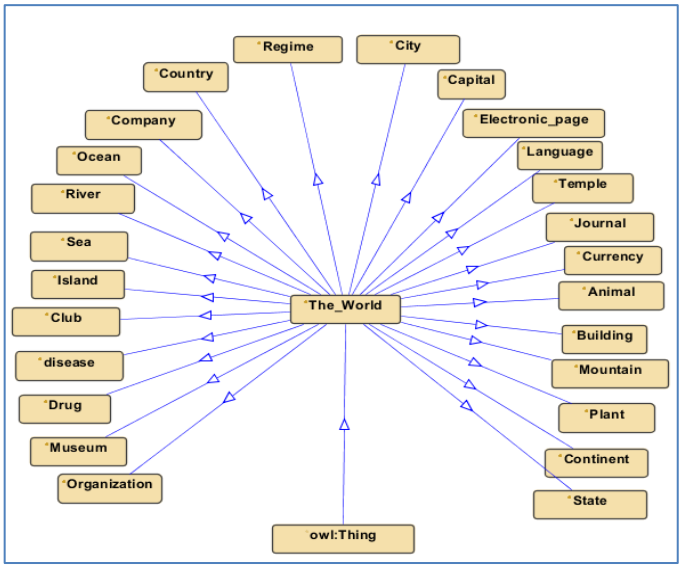

Fig. 3. Top Level Ontology Class.

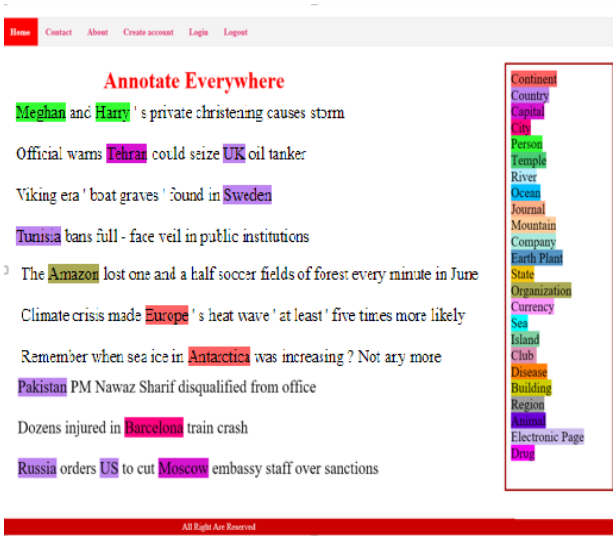

Fig. 4. RSS Final Annotation.

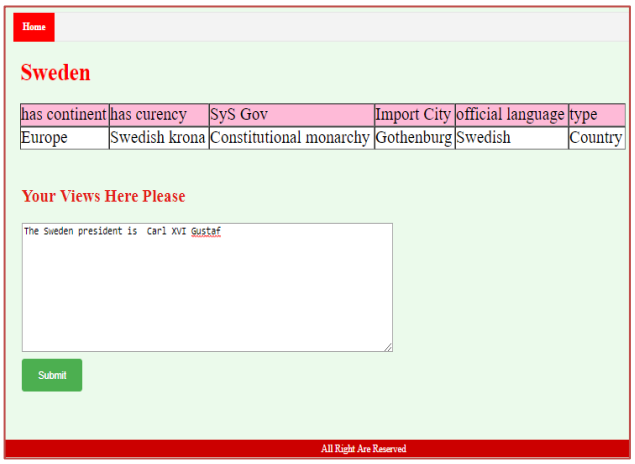

Fig. 5. The Full Annotated Object with Insertion form.

After that, the entered information would be examined and validated in order to enrich the ontology with updated information. If the user needs to know more information about the un-annotated text, the system will search online about the selected topic and treated as an object in the triple store if it is annotated. For example, the user selects an un-annotated word which is "Las Vegas" as shown in "Fig. 6".

Fig. 6. Online Information about un-annotated Objects.

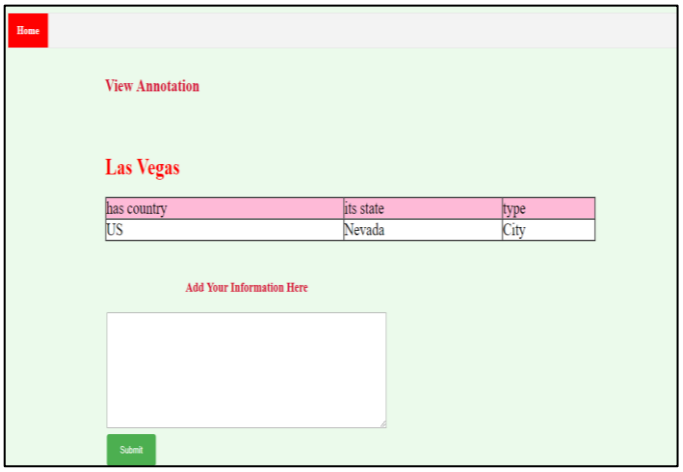

Fig. 7. Enrichment of Ontology with Online Search.

After that, the viewed text on the un-annotated selected object stored in an external database so that it achieve one of two things; one of them, when another user selects the same object; the annotation will be retrieved from the database instead of searching for it once again. The second, If the object belongs to one of the predefined ontology classes; then the system will recommend admin to add this piece of information to the ontology with full description to enrich it. If the user selects it again it will be retrieved directly from the ontology as shown in "Fig. 7".

IV. System Evaluation

The performance of our SNFA tool is our target in this section and how to enhance our annotation tool will be evaluated later on here. To evaluate the performance of updated SNAF semantic tool [11] as we mentioned before, Precision, Recall, and F-measure are calculated based on "equation (1)", "equation (2)" and "equation (3)" [12] [13] [14].

$$\text{recall} = \frac{accurate}{all} \quad (1)$$

$$\text{precision} = \frac{accurate}{accurate + inaccurate} \quad (2)$$

$$\text{F-measure} = 2 * \left(\frac{recall \times precision}{recall + precision}\right) \quad (3)$$

The idea beyond the calculation of F-measure concluded in its ability to reflect accurately the performance of our tool. We applied our experiment on 8 different RSS URLs in the news domain, measured evaluated and compared with results gained from our previous work.

As shown in Table I, values before and after enhancement are presented. The precision almost the same because the same entities retrieved the OWL repository. The average F-measure is improved from 86.9% to 95.5%. This is what we want to reach to achieve the highest efficiency of our system in order to obtain better results every time. On the other hand, "Fig. 8" shows the recall, precision and F-measures variance comparison of our tool before and after enhancement.

TABLE. I. Recall, Precision and F-Measure of SNFA Tool before and after Enhancement

RSS NO.	Old System			New Updated System		
	Precision (%)	*Recall (%)*	*F-measure (%)*	*Precision (%)*	*Recall (%)*	*F-measure (%)*
1	100	90	94.7	100	98	98.9
2	90	70	78.75	100	93	93.6
3	100	82.4	90.4	100	95	97.4
4	95	75	83.8	100	90	94.7
5	100	68.8	81.5	100	80	88.9
6	100	63.6	77.8	100	90	94.7
7	95.7	88	91.7	100	95	97.4
8	90	81.8	85.7	95	90	92.4
Avg			**86.9**			**95.5**

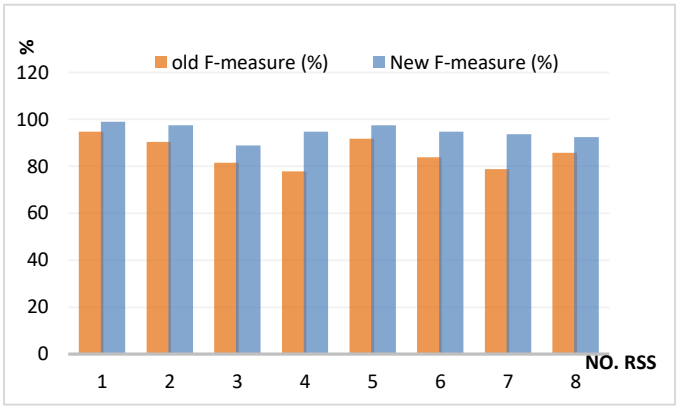

Fig. 8. F-Measure Comparision of SNFA Tool before and after Enhancement.

V. Conclusion and Future Work

Our paper presents an effective framework that allows people to add their annotations. Even if no annotation found, the system will have to search them online and store them in the ontology to be used next time. Our experimental shows that our tool and approach lead to accurate annotations that facilitate information retrieval with high precision and high recall.

We will focus on improving the accuracy of the annotation document by automating the process of extending and enriching the ontology by using machine learning methods in our future work.

References

[1] T.Berners-Lee, M. Fischetti, and M.L. Dertouzos Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor, HarperInformation, 2000, p. 244.

[2] O.Rodríguez Rocha,I.Vagliano and C.Figueroa, "Semantic Annotation and Classification in Practice," IT Professional , vol. 17, no. 2, pp. 33-39, 2015.

[3] "Annotation," [Online]. Available: https://en.wikipedia.org/wiki/. [Accessed 15 May 2019].

[4] F. van Harmelen, "The semantic Web: what, why, how, and when," in IEEE Distributed Systems Online, 2004.

[5] Bontcheva, Kalina, and Hamish Cunningham, "Semantic annotations and retrieval: Manual, semiautomatic, and automatic generation," in Handbook of semantic web technologies, 2011.

[6] Yongxin Liao, Mario Lezoche, Hervé Panetto, and Nacer, "Why, where and how to use semantic annotation for systems interoperability," in 1st UNITE Doctoral Symposium, 2011.

[7] Thomas R.Gruber, "Toward principles for the design of ontologies used for knowledge sharing?," International Journal of Human-Computer Studies, vol. 43, no. 5-6, pp. 907-928, 1995.

[8] "Annotating documents," [Online]. Available: https://www.w3.org/Amaya/User/doc/Annotations.html. [Accessed 24 OCT 2019].

[9] M. P. Kogalovskii, "Semantic Annotating of Text Documents: Basic Concepts and Taxonomic Approach," Automatic Documentation and Mathematical Linguistics, vol. 52, no. 0005-1055, p. 19–26, 2018.

[10] "What is Semantic Annotation?," [Online]. Available: https://www.ontotext.com/knowledgehub/fundamentals/semantic-annotation/. [Accessed 20 10 2019].

[11] Hanaa G.Hendi, Haytham Al-Feel, Ehab E.Hassanein, "Semantic Annotation for News Feeds," in IEEE 8th IEEE Annual InformationTechnology, Electronics and Mobile Communication Conference (IEMCON), Canada, 2017.

[12] K.Holvoet, "What is RSS and how can libraries use it to improve patron service?," Library Hi Tech News, vol. 23, no. 8, pp. 32-33, 2006.

[13] J. Makhou, F.Kubala, R.Schwartz, and Ralph, "Performance Measures For Information Extraction," Proceedings of DARPA broadcast news workshop, 1999.

[14] Powers, D. Martin, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness, and correlation," Journal of Machine Learning Technologies, vol. 2, no. 1, pp. 37-63, 2011.