International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-10 Issue-4, February 2021

Towards an E-journalism Based on Semantic Web Technologies

Ahmed S. Ismail, Haytham Al-Feel, Heba Elbeh, Mohamed Elkawkagy

Abstract: ***The electronic journalism industry became one of the most important achievements of technology in the two decades. Through online media platforms, information and instant news delivered easily and cheaper than before. In addition to that, e-journalism reduces the time and space needed in traditional journalism industry, and hence improve the information lifecycle beginning from collecting reaching to delivering the news to users in convenient ways. On the other hand, Semantic Web technologies enrich the meaning of web content by converting the unstructured data to structured format. So, our proposed works aims to build robust e-journalism system based on Semantic Web technologies to improve the quality of service for journalists and readers.***

Keywords**: *E-journalism, Semantic Web, RDFa, Linked Open Data, SPAR.***

I. INTRODUCTION

Nowadays, all traditional newspapers industry started to publish their content in different ways, one of the most effective publishing techniques is the online press. This achievement helped journalist to share their news and information easily and in a short time compared with its old fashion for delivering it to readers. On the other hand, many of online journals are not indexed or classified into set of specific categories. Also, Reading a huge amount of interrelated information on one page can be very difficult and cause conflict for users who search about a specific piece of information. This interrelation problem forces readers to merge various indirect queries in a manual form before they can get answers to their sophisticated queries. In addition to that, Semantic Web technologies enhance the machine readability of data and then increase the information fusion from different resources to improve the interoperability of web data. Hence, our proposed work is to build an electronic journal portal based on the Semantic Web technologies such as RDFa, Linked Open Data (LoD) and SPARQL query that enables readers to perform different sophisticated queries through a specific web interface. This will increase the integration of different news in different categories that handle the issues of online journalism. The rest of this paper is structured as follow: Section 2 focuses on the related works similar to our objectives. Section 3 overviews the importance of the Semantic Web technologies used. Section 4 presents our proposed architecture. Section 5 shows the results of our experiments. Finally, section 6 concludes our paper and discusses the possible directions for future work.

Manuscript received on February 01, 2021.
Revised Manuscript received on February 08, 2021.
Manuscript published on February 28, 2021.
* Correspondence Author
Ahmed S. Ismail. Computer Science Department, Faculty of Computers and information, Fayoum University, Egypt
Haytham Al-Feel Computer Science Department, Community College, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia
Heba Elbeh Computer Science Department, Community College, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia.
Mohamed Elkawkagy*, Computer Science Department, Community College, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia.

II. LITERATURE REVIEW

Building a robust semantic framework based on journalism and multimedia architecture was discussed in [1]. This architecture depends on XML technologies. Hence they developed an XML Schema or Data Definition Type (DTD) to OWL mapping. The previous mapping was performed via an XML metadata instances to RDF mapping that translates the metadata of XML to the Semantic Web domain. In addition to that, the work in [2] proposed an introduction to the importance of Semantic Web technologies in enhancing the processes of definition, recovery, and exploitation of the online archive of a newspaper. Also, [3] proposed the utilization of semantic technologies and ontologies based on the social behavior such as FOAF and SIOC, to generate newspaper contents. While authors in [4], used Natural Language Generation (NLG) and Big Date in journalism via different platforms such as Hadoop and Spark. In addition to that, the effects of digitalization in traditional journalism industry described in [5]. Authors in [5] proposed a programmable paradigm using a graph theory for motion problems. On the other hand, [6] focused on the impact of the internet and information technologies in China, especially in political and economic sector which had been noticed that it is increasingly affected in journalism studies. A news recommender system was proposed via integrating the Semantic Web technologies for describing and relating news contents and user preferences in order to produce enhanced recommendations [7].

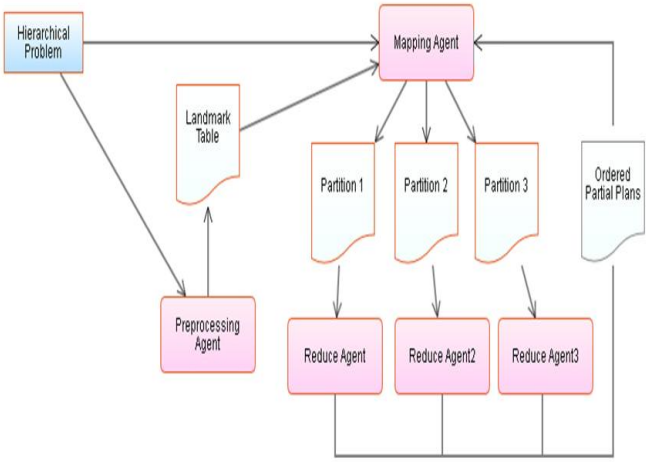

Figure 1: The proposed Architecture

III. BACKGROUND TECHNOLOGIES

In this section we will focus on the Semantic Web technologies which are used through our proposed work such as RDFa, and Linked Open.

3.1 RDFa

Resource Description Framework in Attributes (RDFa) is a specification for attributes that represent a structured data in any markup language [8]. Where the hypertext markup data of structured HTML was reformatted by the RDFa markup tags, hence the publishers didn't need to rewrite specific content in the same HTML document. The core of RDFa is RDF that enables publishers to build their own vocabularies, and then enriching them with maximum interoperability over time. Hence the represented data can be copied and pasted along with its relevant structure.

3.2 Linked Open Data

Linked Open Data is one of the Semantic Web techniques that build typed links between data from different web resources [9]. These may be similar to databases of two organizations in different geographical locations, or simply heterogeneous systems within one organization. In addition to that, Linked Data refers to the published data on the internet in such a way that it is machine-readable in intelligent form. This can be done by defining its meaning explicitly, and linking to other different external resources.

IV. PROPOSED WORK

A semantic news portal proposed in this section based on RDFa tags to enrich the meaning of News content and availability in search engines.In addition to that, we aim to build web interface that facilitate the searching process for readers who look for specific articles or news using SPARQL query language based on the semantic structure of the news portal as shown in Figure 1.

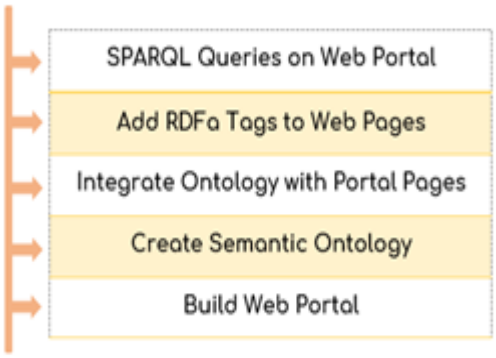

Figure 1: Proposed Architecture

V. IMPLEMENTATION

Through this section, we will show the development process of our proposed system using different web development languages and tools such as php web language and MySQL database to build our Newspaper portal. RDFa was used to enrich news pages where authors could write their news article semantically structured in Resource Description Framework (RDF) structure. On the other hand, Linked open Data (LoD) used in the process of linking different news resources for a specific topic which enables users to find everything about a topic they search for. Finally, A SPARQL endpoint attached with our web portal was implemented to allow users asking for sophisticated queries as shown in Figure.2 and Figure.3.

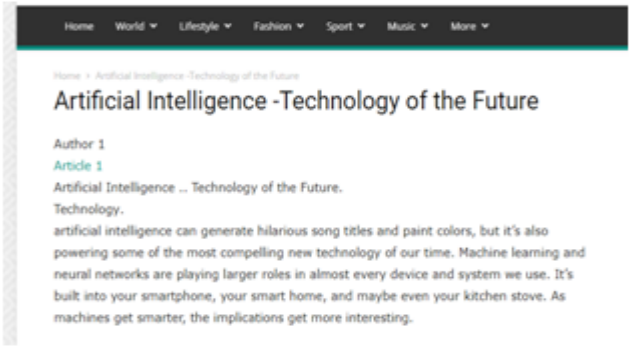

Figure 2: NEW webpage using RDfa code (backend)

VI. RESULTS

This section represents the conducted experiments in our system, in addition to the evaluation criteria of the proposed architecture as shown in Table 1, using Gerber's evaluation method to clarify the weakness of this architecture and focus on the possible modifications that can be performed to our new architecture.

```
SELECT ?content ?Author
WHERE {?Author NEWS:author_of NEWS:Article_1 . NEWS:Article_1 NEWS:Contains ?content }
```

Figure 3: SPARQL query for retrieving sophisticated queries for our proposed system

Table 1: Evaluation Criteria for our architecture

Criteria	Conformity
Clearly defined context	Conformed: The architecture components are languages required for implementing our architecture that matches with Gerber results
Hiding of implementation details	Not Conformed: There is a lot of Implementation details in the architecture
Clearly defined functional layer	Partially Conformed: The functions of some layers are clearly known from their names, while others may have technological names which do not explain the functionality clearly.
Appropriate level of abstraction	Not Conformed: our system cannot be abstracted as one component because it consists of various technologies & functional layers

The evaluation method conducted by Gerber depends on a number of criteria which integrated from architecture design and software engineering as we described below [10].

- Clearly defined context: the criterion by which we could clarify the components required in this architecture and the reason for their integration [10]
- Appropriate level of abstraction: the factor that evaluates what extent the system can be viewed as one component[10-13].
- Hiding of implementation details: the factor that evaluate the quality of our system design where hiding the implementations details from the architectural model is a good sign for the quality of system.
- Clearly defined functional layers: the factor that evaluates the function of each component.
- According to the conducted evaluation for our architecture, it appears that this architecture seems good enough to represent the robustness of our system especially after applying the availability, Simplicity, and security factors to our evaluation method where .
- Availability is the degree to which a system is in a specified operable and committable state at any time it is required to be used.
- Simplicity is having fewer tools to achieve the same results effectively
- Security is a group of threshold restrictions that must be fit the system to be evaluated and deemed acceptable.

Our experiments were conducted on set of news webpages to measure the reaching of these news easily and quickly for users using RDFa and without it. Precision which is the fraction of relevant instances among the retrieved instances was measured according to the equation.1[13] while recall which is the fraction of relevant instances that have been retrieved over the total amount of relevant instances was measured according to the equation.2[13] based in the confusion matrix [14] that classify the performance of a classification on a set of experimental data for which the true values are known and the negative values are false as shown in Figure.4.

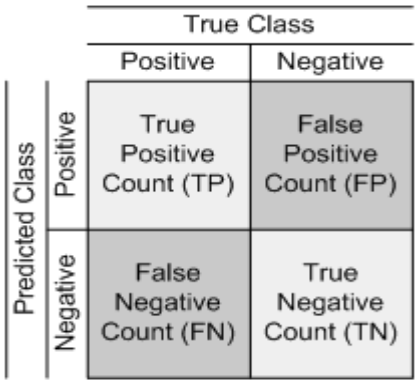

Figure 4: Confusion Matrix

According to that, the following table illustrates the value of precision, recall and F-Score as a measurement method to calculate the accuracy. The F-score[13] is an average of both precision and recall, where an F-score reaches its best value at 1 and worst at 0.as shown in equation.3

$$PRECISION(A) = \frac{TP}{TP + FP_A} \quad (1)$$

$$RECALL(A) = \frac{TP_A}{\text{Total Positives}} \quad (2)$$

$$F1\ Score = 2 * \frac{precision * recall}{precision + recall} \quad 3$$

On the other hand, our experiments were performed in two different ways, one was for 5 articles news provided with RDFa and another 5 articles news without RDFa in their content. Table 2 shows the precision, Recall and F-Score for web pages in RDF syntax, While Table 3 shows them in web pages without RDFa as shown in Figure 5.

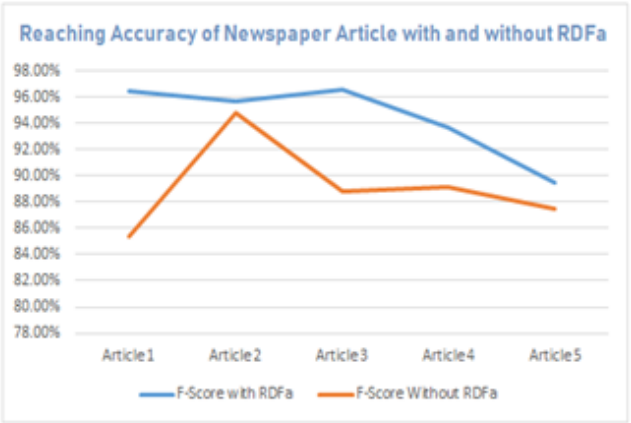

Figure 5: Difference between Accuracy of the same articles in two cases: case 1: With RDFa || Case 2: Without RDFa

Table 2: Table of Precision, Recall and F-score of Web pages provided with (RDFa syntax)

Article + RDFa	Precision	Recall	F-Score
Article 1	100%	93.30%	96.50%
Article 2	100%	91.80%	95.70%
Article 3	100%	93.50%	96.60%
Article 4	98.90%	89.20%	93.70%
Article 5	99.40%	81.40%	89.50%

Table 3: Table of Precision, Recall and F-score of Web pages without (RDFa syntax)

Article - RDFa	Precision	Recall	F-Score
Article 1	90%	81.20%	85.40%
Article 2	89%	80%	94.80%
Article 3	93%	85%	88.80%
Article 4	92%	86.50%	89.20%
Article 5	93%	82.60%	87.50%

VII. CONCLUSION

Through this research, we proposed a new architecture for e-newspaper based on semantic web technologies. Our proposed architecture based on RDFa, Linked Open Data and SPARQL query language to provide the Newspaper web portal with more meaningful content and structuring format. This achievement helps authors and readers of newspaper articles, where the author could write his article in structured format and classify each one of them into its appropriate category. This will improve the search engine optimization of the web portal in search engine indexing. On the other hand, it helps reader to find his intended articles quickly as he wants. Moreover, our system provides users with query endpoint to run sophisticated queries on related news from different categories based on SPARQL and Linked Open data. Eventually, we plan to improve our architecture to cover different languages and then build a recommender system that helps readers to select the topic related to their interests.

REFERENCES

1. R. García, P. Ferran and G. and Rosa, ""Ontological infrastructure for a semantic newspaper."," Semantic Web Annotations for Multimedia Workshop, SWAMM., 2006..
2. P. Castells, F. Perdrix, E. Pulido and M. Rico, "Semantic Web Technologies for a Digital Newspaper Archive," in European Semantic Web Symposium, May 2004.
3. B. R. Heravi, B. Marie and B. and John, ""Towards Social Semantic Journalism."," in In Sixth International AAAI Conference on Weblogs and Social Media., 2012.
4. K. N. Dörr, ""Mapping the field of algorithmic journalism."," in Digital journalism, (2016).
5. M. G and T. and A. N, "THE IMPACT OF DIGITALIZATION OF NETWORK SPACE ON JOURNALISM EDUCATION."," in Медиаобразование, (2019).
6. T. Xu, How has the Internet Impacted on Traditional Journalism in the Context of China?, May 2015.
7. C. I, "News@ hand: A semantic web approach to recommending news," in Adaptive hypermedia and adaptive web-based systems, 2008.
8. R. Meusel, P. Petar and B. and Christian, ""The webdatacommons microdata, rdfa and microformat dataset series."," in In International Semantic Web Conference,, 2014.
9. A. Mikroyannidis, D. John, M. Maria, N. Barry and S. and Elena, ""Teaching linked open data using open educational resources."," in In Open Data for Education,, 2016.
10. W. S. Eller, J. G. Brian and R. and Scott E, Public administration research methods: tools for evaluation and evidence-based practice., Routledge, 2018.
11. A.Abdelatey, M. Elkawkagy, A. El-Sisi, and A. Keshk, "A multilateral agent-based service level agreement negotiation framework". Published in International Conference on Advanced Intelligent Systems and Informatics. Springer 2016-pages 576-586
12. A Abdelatey, M Elkawkagy, A El-Sisi, A Keshk, "A Repaired Genetic Algorithm-based Approach for Web Service Security Negotiation", published in International Conference on Computer Theory and Applications, 2016.
13. F. Morstatter, W. Liang, N. Tahora H, C. Kathleen M and L. and Huan, ""A new approach to bot detection: striking the balance between precision and recall."," in IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2016.
14. K. M. Ting, ""Confusion matrix."," in " Encyclopedia of Machine Learning and Data Mining, 2017.

AUTHOR PROFILE

Ahmed Salama is a lecturer at the Information Systems Department, Faculty of Computers & Information, Fayoum University, Egypt, And October 2016- Currently. He got his MSc in 2016 from the Faculty of Computers & Information, Cairo University in 2016. He got his Ph.D. degree from the Faculty of Computers & Information, Fayoum University in 2020. A co-founder of the Arabic Chapter of DBpedia with cooperation with the DBpedia Project at Leipzig University in 2016. He published more than 6 papers till now in different fields in computer technologies such as the semantic web and Big data.

Haytham Al-Feel is a Full Professor in Data Science. He is a staff member at the Computer Sciences department, Community College, Imam Abdulrahman Bin Faisal University, KSA from October 2018 till now. He got his PhD in 2009 from the Faculty of Electronics Engineering, Menofia University in 2009. He was a Post-Doc at the Semantic Web Cooperate Research Group, Freie University Berlin, Germany between 2011 and 2012. Prof.Haytham has published more than 35 paper in international journals and conferences.

Heba Elbeh is Lecturer in Artificial intelligence. She is a staff member at the Computer Sciences department, Community College, Imam Abdulrahman Bin Faisal University, KSA from November 2016 till now. He got his PhD in 2012 from the Faculty of Engineering, Ulm University, Germany. Her research focuses on Artificial In intelligence, Block chain, and Big data.

Mohamed Elkawkagy is Associate Professor in Artificial intelligence. He is a staff member at the Computer Sciences department, Community College, Imam Abdulrahman Bin Faisal University, KSA from November 2016 till now. He got his PhD in 2011 from the Faculty of Engineering, Ulm University, Germany. He was a Post-Doc at the Artificial Intelligence Research Group, Ulm University, Germany between 2011 and 2013. His research focuses on Artificial In intelligence, cloud computing, and Big data.