Research on Using Semantic Web Technologies in Content Management Systems

Tsvetanka Georgieva-Trifonova, Svetlin Marinov

"St. Cyril and St. Methodius" University of Veliko Tarnovo, Veliko Tarnovo, Bulgaria

Abstract **– In the present paper, the usage of Semantic web technologies for the purposes of Content Management Systems (CMSs) is explored. A review of the research related to semantic CMS is performed. The distinguishing features between CMSs that support Semantic web technologies and native semantic CMSs are discussed. The possibilities of semantic CMSs to overcome the challenges posed by big data are considered.**

Keywords **– Semantic web technologies, content management systems, big data.**

1. Introduction

Web content management systems facilitate content maintenance activities and have gained growing popularity due to their flexibility, accessibility to a wide range of users, adaptability in developing applications in different domains [17]. Semantic web technologies allow expanding the capabilities of traditional CMS, whose main purpose is to create, publish, edit and manage content. This leads to the emergence of the so-called semantic CMS.

Semantic CMSs provide semantic functionalities that include data storage and processing from semantic services, knowledge extraction from data, semantic reasoning.

DOI: 10.18421/TEM114-46
https://doi.org/10.18421/TEM114-46

Corresponding author: Tsvetanka Georgieva-Trifonova, *University of Veliko Tarnovo, Bulgaria.*
Email: cv.georgieva@live.uni-vt.bg

Received: 10 August 2022.
Revised: 30 September 2022.
Accepted: 10 October 2022.
Published: 25 November 2022.

For this purpose, an approach is required to processing the data separately from the page/document. Therefore, semantic CMSs differ from traditional CMSs in their ability to interact with content, as well as automatically store, extract and manage semantic content metadata [1], [12].

The present study is motivated by the need to mark progress in the development of semantic CMSs. For this aim, the scientific research related to semantic CMS are investigated. A distinction is made between CMSs that support semantic web technologies and native semantic CMSs, and their possibilities for dealing with the challenges posed by big data are examined.

2. Related Work

The content management systems are extensively used by news and media organizations, e-commerce websites, digital libraries, the television and movie industry, and educational institutions for efficient content management. For the purposes of the Interactive Knowledge Stack research project:

- Laleci et al. (2010) [16] describe how to improve the semantic capabilities of a CMS by building on the already available semantics in content models and turning them into ontologies to allow the use of implicit relationships between content elements for sophisticated content search and navigation mechanisms;
- In [6], an architecture for a semantic CMS is represented and its value is validated through two implementations;
- In [1], the authors' experience in the development and construction of semantic components and their integration into existing CMSs is described.

Some studies focus on specific domains:

- Media content management;

In [10], a Rhizomer content management system based on semantic web technologies is described. All content is described using semantic metadata semi-automatically extracted from multimedia content. As an application scenario of the platform, its use in a media company is described, where it manages audio

content and semantically annotates its speech transcript.

- The renewable energy sector [18];
- Improving the performance of learning management systems;

An ontology-based learning content management system is proposed in [22]. The ontologies are created to describe the learner profile and learning content, which are used to classify learning material, organize course topics, and store the information related to the learner profile, and therefor they are useful to the user in searching for a more accurate and appropriate learning object.

Bratsas et al. (2012) [3] represent the results of a study involving the customization and implementation of MediaWiki and Drupal to evaluate their potential for exposing university learning resources to the Linked Data Web. For this purpose, an in-depth comparison of the possibilities of these platforms to function as learning management systems is conducted.

Cardoso (2017) [5] demonstrates the applicability and advantages of using Semantic web technologies by developing a real application that is a system for management of semantic university course, which is entirely based on the Semantic web and uses the latest technologies in the field such as OWL, RQL (*Resource Query Language*), SWRL (*Semantic Web Rule Language*).

- E-gov applications;

In [4], a semantic content management system capable of handling multimedia content designed to support e-gov applications is considered. All information is described using semantic metadata modelled and represented using RDF (*Resource Description Language*) / OWL (*Web Ontology Language*).

- Web information portal [19];
- Journalistic writing;

Silva de Deus et al. (2018) [21] represent a CMS prototype that focuses on building semantic annotations based on a domain ontology, reusing annotations in search, and building relationships between stored texts.

Vogt et al. (2019) [23] propose a framework for the development of FAIR (findable, accessible, interoperable, reusable) system for management of Semantic web content, called SOCCOMAS (*Semantic Ontology-Controlled application for web COntent MAnagement Systems*).

The successful implementation of the idea of the Semantic web [2] requires the development of beneficial applications, based on applying the Semantic web technologies, which can be greatly assisted by the availability of semantic CMSs. The present paper systematizes the characteristics of the two categories of semantic CMSs – those that support Semantic web technologies and native semantic CMSs. Their capabilities to address big data challenges are discussed.

3. Types of Semantic Content Management Systems

Considering the advantages that Semantic web technologies propose, existing CMSs are upgrading their new versions by providing capabilities to work with these technologies. On the other hand, CMSs can currently be found that are entirely based on Semantic web technologies to represent, store, search content. From this point of view, we distinguish two types of semantic CMSs, which are discussed in more detail in the present section.

3.1. Cmss that Support Semantic Web Technologies

CMSs that support Semantic web technologies mainly offer plugins, extensions about (Fig. 1):

- Data representation in RDF model [26];

This capability is intended to facilitate the process of transforming data into an RDF model that can be performed by CMS users even without significant experience in RDF modeling.

- Possibility for processing Microdata [27] and/or RDFa [28];

HTML (*HyperText Markup Language*) Microdata is an extension of HTML5 [29] designed to embed machine-readable data in web documents through the usage of attributes. RDFa (*Resource Description Framework in Attributes*) proposes a set of markup attributes intended for augmenting the information available at the Web by adding machine-readable hints. Unlike HTML Microdata, RDFa is applicable to any markup language. The role of a CMS is to apply the processing rules described in the specifications to extract RDF from an HTML document.

- Execution of SPARQL (*SPARQL Protocol and RDF Query Language*) queries [25].

The ability to select, filter, sort, precisely search the RDF data is implemented by executing queries in the SPARQL language. Representing the data in an RDF model using Semantic web technologies limits the search result due to the presence of a semantic description about the stored data retrieved from the web pages.

These functionalities could be available through additional plugins, extensions in popular CMSs such as Drupal [7], [13], [15]; MediaWiki, Wordpress, etc. Their usage could significantly increase the amount and thematic range of information represented through Semantic web technologies, which in turn would redound to the implementation of the Semantic web idea.

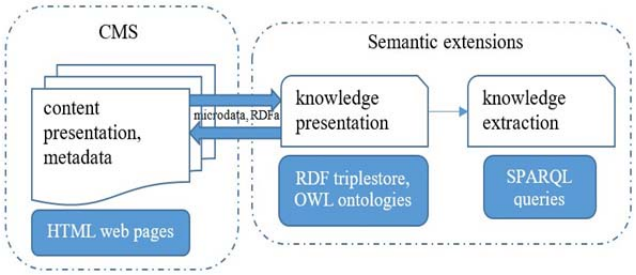

Figure 1. CMSs that support Semantic web technologies

3.2. Native semantic CMSs

The native semantic CMSs implement the functionalities analogous to the traditional ones, but using the standard W3C models defined for the purposes of the Semantic web and the corresponding data representation formats using Semantic web technologies as a base. They are established on the standards for Semantic web – RDF and OWL [24], for representing knowledge and publishing it with correspondence to the principles of linked data on the Web [14]. Therefore, their main functionalities include (Fig. 2):

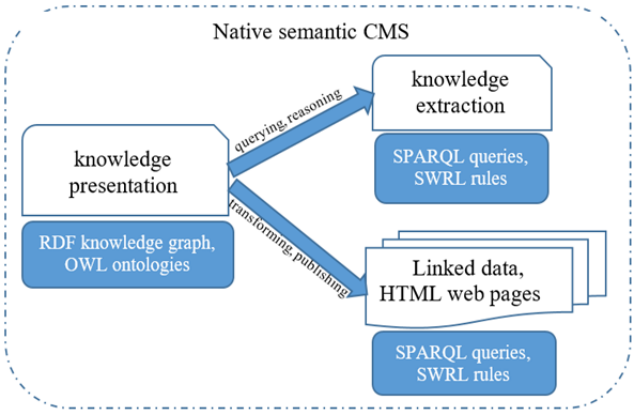

Figure 2. Native semantic CMSs

- Creation of ontologies represented through OWL – both user-defined ontologies and reuse of existing ontologies for specific domains;
- Description of metadata using standard vocabularies such as Dublin Core, FOAF, SIOC, SKOS, etc.;
  - ❖ Semantic search;
  - ❖ Semantic reasoning;
  - ❖ Publishing Linked data.

A representative of the native semantic CMSs that has gained popularity is GNOSS, (https://www.gnoss.com/en/semantic-framework/semantic-content-management).

Another similar solution is proposed by Ximdex, (https://www.ximdex.com). Ximdex is a semantic CMS that uses XML (*eXtensible Markup Language*) as the main format for data representation. It offers a convenient WYSIWYG editor for XML documents. Structured XML documents are automatically transformed via XLST (*eXtensible Stylesheet Language Transformations*) templates into formats needed to generate HTML content, RDF models and applications (such as JSON, etc.). The advantage of the XML representation of the data and its metadata descriptions is that the transformation can be based on the meaning embedded in the data.

In addition, Ximdex provides:

- Configurable search engine that can simultaneously work with semantic documents (RDF), structured (XML) and unstructured documents;
- Tools for managing visual semantic tags (Xtags component) that include:
- An ontology browser for common ontologies (such as schema.org) or user-defined ontologies for structured tags;
- Automatic tag suggestions (using DBpedia, freebase) or other tags in the repository;
- User-defined tags.
- Advanced dynamic semantic publishing of RDF, OWL, etc.
- Linked open data management module;
- API REST with semantic capabilities to provide an endpoint for linked open data.

3.3. Semantic CMSs and big data

An essential challenge emerges from the need to find the most appropriate information in a large dataset [8], [9]. Information systems must deal with the efficient management of heterogeneous content that is stored in distributed repositories and poses the characteristics of the big data [20]. CMS must provide higher quality of data management, since the amount of data is expected to continue to grow. This could be accomplished by supporting the Semantic web technologies and especially those for building ontologies.

The purpose of ontology is to provide a common understanding of a domain that can be shared, reused, and exchanged between heterogeneous and distributed systems. The very purpose of ontologies implies an opportunity to overcome problems related to the main distinguishing characteristics of big data – variety, veracity, value, variability, which was confirmed by the conducted scientific review of the existing research experience in [11]. As a result, modeling and building ontologies in various domains support the management and analysis of big data. Using a native semantic CMS, in turn, could significantly facilitate and support this process.

On the other hand, a challenge for a native semantic CMS is to ensure scalability in terms of data size, i.e. dealing with the volume characteristic of big data. In addition, such a semantic CMS could provide data processing efficiency at data size enlargement, based on the implementation of

technologies that allow the implementation of data sharding, the execution of distributed queries. These technologies are undergoing substantial development with the growing popularity of non-relational databases for distributed systems purposes.

4. Conclusion

The present study is addressed on the application of the Semantic web technologies for the task implementation in a content management system. A survey of the research related to semantic CMSs is done. Two types of semantic CMS are distinguished according to the way they include Semantic web technologies in implementing their functionalities. The challenges produced by big data and the possibilities of the semantic CMSs to overcome them are discussed.

CMSs that support Semantic web technologies extend the capabilities of existing traditional CMSs, making them more widely applicable. The users who already apply them have the option to add a new plugin or extension to their installation. The native semantic CMSs need time to promote and establish themselves among new users. On the other hand, they provide full support for Semantic web technologies in accordance with W3C recommendations, built-in implementation of semantic services.

References

[1]. Behrend, W., & Damjanovic, V. (2013). Developing Semantic CMS Applications-The IKS Handbook. *Salzburg: Salzburg Research Forschungsgesellschaft*.

[2]. Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific american*, *284*(5), 34-43.

[3]. Bratsas, C., Dimou, A., Antoniou, I., Ioannidis, L., & Bamidis, P. (2012). Semantic CMS and wikis as platforms for linked learning. In *CEUR Workshop Proceedings* (Vol. 840, No. IKEECONF-2019-238). Aristotle University of Thessaloniki.

[4]. Cappetta, D., D'Elena, S., Moscato, V., Orabona, V., Palmieri, R., & Picariello, A. (2014, August). A Semantic Content Management System for e-Gov Applications. In *DATA* (pp. 440-445). doi: 10.5220/0005146404400445

[5]. Cardoso, J. (2007). Developing course management systems using Semantic Web technologies. *The semantic web: real-world applications from industry*.

[6]. Christ, F., & Nagel, B. (2011, September). A Reference Architecture for Semantic Content Management Systems. In *Gesellschaft für Informatik eV (GI) publishes this series in order to make available to a broad public recent findings in informatics (ie computer science and informa-tion systems), to document conferences that are organized in co-operation with GI and to publish the annual GI Award dissertation.* (pp. 135-148).

[7]. Corlosquet, S., Cyganiak, R., Polleres, A., & Decker, S. (2009). RDFa in Drupal: Bringing Cheese to the Web of Data. In *SFSW@ ESWC*.

[8]. Fernández, J. D., Arias, M., Martınez-Prieto, M. A., & Gutiérrez, C. (2013). Management of big semantic data. *Big data computing*, 131-167. doi: 10.1201/b16014-7

[9]. Fisher, M., & Sheth, A. P. (2005). Semantic Enterprise Content Management. *The Practical Handbook of Internet Computing*, *2*, 9-21. doi: 10.1201/9780203507223.ch9

[10]. García, R., Gimeno, J. M., Perdrix, F., Gil, R., & Oliva, M. (2008, September). The rhizomer semantic content management system. In *World Summit on Knowledge Society* (pp. 385-394). Springer, Berlin, Heidelberg. doi: 10.1007/978-3-540-87781-3_42.

[11]. Georgieva-Trifonova, T., & Galabov, M. (2021). Semantic Web Technologies for Big Data Modeling from Analytics Perspective: A Systematic Literature Review. *Baltic Journal of Modern Computing*, *9*(4), 377-402. doi: 10.22364/bjmc.2021.9.4.01

[12]. Gönül, S., & Sinaci, A. A. (2012, September). Semantic content management and integration with JCR/CMIS compliant content repositories. In *Proceedings of the 8th International Conference on Semantic Systems* (pp. 181-184).

[13]. Havlik, D. (2011, June). Building Environmental Semantic Web Applications with Drupal. In *International Symposium on Environmental Software Systems* (pp. 385-397). Springer, Berlin, Heidelberg. doi: 10.1007/978-3-642-22285-6_42.

[14]. Heath, T., & Bizer, C. (2011). Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, *1*(1), 1-136. https://doi.org/10.1007/978-3-031-79432-2_2

[15]. Kalou, A. K., Koutsomitropoulos, D. A., & Solomou, G. D. (2013, November). CMSs, Linked Data and Semantics: A Linked Data Mashup over Drupal for Personalized Search. In *Research Conference on Metadata and Semantic Research 390,* 48-59. Springer, Cham. doi: 10.1007/978-3-319-03437-9_6.

[16]. Laleci, G. B., Aluc, G., Dogac, A., Sinaci, A., Kilic, O., & Tuncer, F. (2010). A semantic backend for content management systems. *Knowledge-based systems*, *23*(8), 832-843.

[17]. Martinez-Caro, J. M., Aledo-Hernandez, A. J., Guillen-Perez, A., Sanchez-Iborra, R., & Cano, M. D. (2018). A comparative study of web content management systems. *Information*, *9*(2), 27. DOI: 10.3390/info9020027.

[18]. Ngonga Ngomo, A.-C., Heino, N., Lyko, K., Speck, R., & Kaltenbock, M. (2011). SCMS – Semantifying Content Management Systems, Lecture Notes in Computer Science, vol. 7032. Springer, Berlin, Heidelberg, pp. 189–204. doi: 10.1007/978-3-642-25093-4_13.

[19]. Orabona, V., Palmieri, R., Moscato, V., Picariello, A., D'Elena, S., & Cappetta, D. (2015, July). An Application of Semantic Web Technologies to Enhance Content Management in Web Information Portals. In *DATA* (pp. 378-381). doi: 10.5220/0005579803780381.

[20]. Priyadarshini, R., & Tamilselvan, L. (2014). Document based semantic CMS in cloud. *Information Technology Journal*, *13*(2), 217-230. DOI: 10.3923/itj.2014.217.230

[21]. de Deus, V. S., Ishikawa, E., Oliveira, E. C., Victorino, M., Neto, B. M., Groenli, T. M., & Ghinea, G. (2018, September). Towards a semantic-based content management system for journalistic writing. In *Proceedings of the 10th International Conference on Management of Digital Ecosystems* (pp. 141-148). doi: 10.1145/3281375.3281407

[22]. Thakar, U., Meena, A., & Meena, A. (2010). OLearner–An Ontology Based Learning Content Management System to Support Semantic Search and Contribution of Learning Objects. *Journal of Algorithms & Computational Technology*, *4*(4), 587-605.

[23]. Vogt, L., Baum, R., Bhatty, P., Köhler, C., Meid, S., Quast, B., & Grobe, P. (2019). SOCCOMAS: a FAIR web content management system that uses knowledge graphs and that is based on semantic programming. *Database*, *2019*, 1-22. doi: 10.1093/database/baz067

[24]. W3C. (2012). OWL 2 Web Ontology Language. Retrieved from: https://www.w3.org/TR/owl2-primer/ [accessed: 20 May 2022].

[25]. W3C. (2013). SPARQL 1.1 Query Language. Retrieved from: https://www.w3.org/TR/sparql11-query [accessed: 25 May 2022].

[26]. W3C. (2014). RDF 1.1 Concepts and Abstract Syntax. Retrieved from: https://www.w3.org/TR/rdf11-concepts [accessed: 25 May 2022].

[27]. W3C. (2014). Microdata to RDF – Second Edition: Transformation from HTML+Microdata to RDF. Retrieved from: https://www.w3.org/TR/microdata-rdf/ [accessed: 30 May 2022].

[28]. W3C. (2015). RDFa 1.1 Primer - Third Edition Rich Structured Data Markup for Web Documents. Retrieved from: https://www.w3.org/TR/rdfa-primer/ [accessed: 17 June 2022].

[29]. WHATWG. (2022). HTML: Living Standard. Retrieved from: https://html.spec.whatwg.org/ [accessed: 21 June 2022].