

Creation, Storage and Presentation of Information Content – Semantics, Sharing, Presentation, and Archiving

Jan Masner, Pavel Šimek, Eva Kánská, Jiří Vaněk

Department of Information Technologies, Czech University of Life Sciences Prague, Czech Republic

Abstract

People are getting more and more used to consume a digital, online content. Many outlets switched to online publication or at least increased their online presence. Besides, online publication is not the only domain of publishing houses. Many different organisations and companies - including those from area of agriculture and rural development - provide online content in form of articles. Importance of semantic web is growing constantly. Together with metadata descriptions, it is necessary for all the current search engines, smart assistants and AI technologies. Public standards and open source software can significantly speed up development and reduce costs when it comes to the Internet and World Wide Web. The paper provides overview of an updated methodology for creation, storage and presentation of online information content in World Wide Web environment. The latest research was focused mainly on presentation and semantics. The whole research process is established as well as the final formulation of the methodology.

Keywords

WWW, semantics, semantic web, metadata, information content, CMS, HTML, WYSIWYG, Web 4.0.

Masner, J., Šimek, P., Kánská, E. and Vaněk, J. (2019) "Creation, Storage and Presentation of Information Content – Semantics, Sharing, Presentation, and Archiving", *AGRIS on-line Papers in Economics and Informatics*, Vol. , No. 1, pp. 75-82. ISSN 1804-1930. DOI 10.7160/aol.2019.110108.

Introduction

The paper builds on the previous work published by (Masner et al., 2018). The paper provides an updated version of the methodology for Creation, Storage and Presentation of Information Content, newly referred as WICM (Web Information Content Management). The paper provides overview of the research design and all the methods used.

We have seen a significant departure from classical printed media such as newspapers, magazines, and even books for more than twenty years. People are getting more and more used to consume a digital (online) content. In order to stay relevant in light of these changes, many outlets switched to online publication or at least increased their online presence (Das et al., 2009). Besides, online publication is not the only domain of publishing houses. Many different organisations and companies provide online content in form of articles. Establishing and managing websites and portals have become easier as of late. Due to the progressive development of internet and web technologies, especially Content Management Systems (CMS), even users without knowledge of the desired technologies

such as HTML, CSS, and JavaScript can manage the online content nowadays (Brown, 2014).

Current information content should focus not only on the appearance of the result shown in a browser but also to be accessible for humans as well as for machines (Minin et al., 2015). The content is also commonly viewed on variety of devices, especially smartphones and tablets (Šimek et al., 2014). The structuring of content then helps to display it more responsively. Besides, structuring and metadata are essential in open data publication (Stoček et. al, 2018).

As (Rudman and Bruwe, 2016) stated: "Web 3.0 entails an integrated Web experience where the machine will be able to understand and catalogue data in a manner similar to humans." So, Web 3.0 is usually considered by authors as semantic web. Furthermore, there are talks about "Web 4.0". There is not a unified definition of the term yet. Some authors define it as a symbiotic web with interactions between humans and machines (Aghaei, 2012; Choudhury, 2014), many stated that there will be an important role of IoT technologies (Nath and Iswary, 2016;

Srirama, 2017), and some mention the important role of Big Data technologies (Peinl, 2016). Taking everything into account, until there is a real artificial intelligence which understands human language, the semantics on the web is crucial. Semantics is necessary for all the current smart assistants such as Google Home, Amazon Echo, Siri (Sweeney, 2016).

Open source software and open source CMSs in particular are highly exploited in the field of online publication. Not only in areas such as agricultural sector, rural development, government and local government and non-profit sector, the open source software is very popular. It is a logical way of saving expenses for own development (Šimek, et. al., 2017).

Use of WYSIWYG (What You See Is What You Get) editors for content creation and update is very widespread. These editors strive to work like conventional desktop text editors such as Microsoft Word. Many users use the Word application to create documents for printing. Nevertheless, WWW environment is different and has many specifics. Therefore, the output is not always as intended and seen in the editor. The WYSIWYG editors are not able to simulate different devices and responsiveness. Moreover, as (Khalili et al., 2012; Khalili and Auer, 2015) state, there is lack of support of semantics. On the other side, advanced ways of content creation, such as composing

from separate, simpler and more independent blocks using some kind of blocks are becoming more widespread (Nikolic and Šilc, 2016), (Czerniak, 2015).

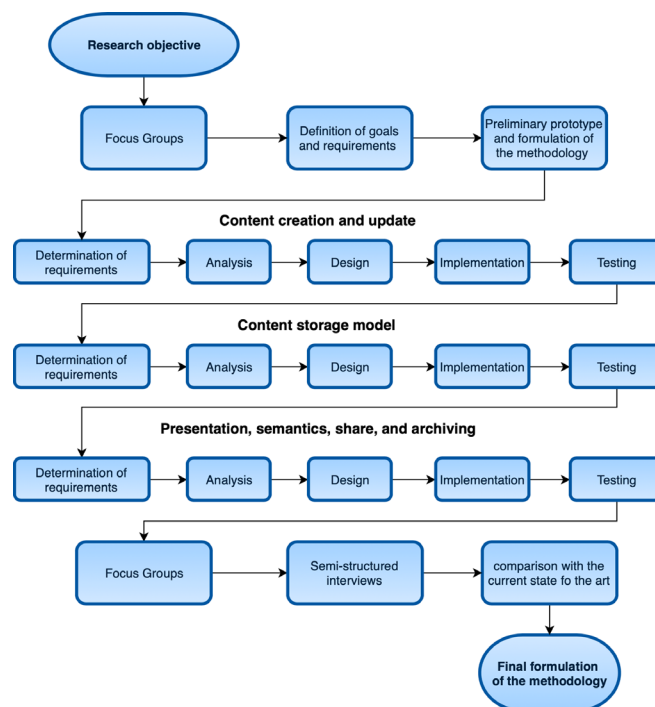
Current trend in web development leans towards small applications modules, reusable components and microservices. As (Sahay, 2003) shows, the standardisation simplifies software development and enables utilisation of widely used libraries, modules and frameworks based on open source code. Using open source software components has therefore a positive influence on development speed (Merilinn and Matinlassi, 2006).

The main objective of the proposed WICM methodology is to bring standardisation to creation, storage and presentation of the information content in WWW environment. The methodology was divided into 3 main modules. Each module deals with different area of interest. The modules are as follows:

- Creation and update
- Storage model
- Presentation, semantics, share and archiving

Materials and methods

The whole research process utilizes several methods as shown in Figure 1. The methodology development was based on prototyping approach



Source: own processing

Figure 1: Research process design.

according to (Bally et al., 1977). It has been divided into three main stages according to the three defined modules. First of all, we developed a preliminary software prototype. After that, the Focus Group session was held. The prototype was shown to the participants. The session was mainly dedicated to a collection of users' needs. After that, the detailed objectives and requirements for the methodology has been defined. The previous paper dealt mostly with the Content creation and update and the Content storage model modules.

The objectives and requirements for the methodology have been updated, based on the preliminary and follow-up research. According to the modules, they are as follows:

- Creation and update
 - Users without knowledge of internet technologies should be able to create and update the content
 - The process of creation and update should not allow to make inconsistent output (in terms of HTML code)
 - Data input should be user friendly without any need for training
 - Users should be able to see the result of the inserted content
- Storage model
 - There is a need for general model of storage without regards to database technology
 - The structure should respect other modules
- Presentation, semantics, share and archiving
 - The stored content should not contain any CSS code
 - The structure of the stored content should be transformable for use in various devices such as tablets and smartphones
 - There should be a way to define presentation rules, especially HTML structure
 - The stored content should be resistant to redesigns
 - The methodology should follow semantics rules
 - The storage model should allow metadata enrichment
- The methodology should follow best practices for SEO

- There should be standardised data format and structure for content transfer, share, exchange and backup

Within the third stage, an analysis of semantic web, metadata description, and data exchange formats has been conducted. In WWW environment, there are basically two formats used to store and exchange data – XML and JSON. As (Sandrih et al., 2017) stated, both formats are very commonly used and are exchangeable and inter-transformable. Many authors point out that JSON format is faster in processing and less data-intensive, especially for use with mobile devices (Jorstad et al., 2008; Nurseitov et al., 2009; Lin et al., 2012). JSON is therefore appropriate exchange format to be used within web services and by server-side APIs in communication with mobile applications. To facilitate portability of the content, an exchange format based on the storage structure needs to be defined.

After the third stage, the Focus Groups method was used again. The same group of participants as previously was involved. The developed prototype was introduced and discussed. The session was mainly focused on the user-related part of the proposed methodology. The prototype is shown in Figure 2 and Figure 3.

The methodology and its components were discussed with several web professionals and developers. Semi-structured interviews were conducted. The interviews were focused on the other side of the methodology – storage model, technological presentation in browsers and aspects related to developers and administrators. Guide for the semi-structured interviews included following questions:

- Can you evaluate the use of content structuring for content creators?
- How do you evaluate the storage model? Does it contain all the necessary fields?
- Can the proposed standardisation help speed-up the development process?
- How are you using metadata description? Is the proposed method beneficial?
- Have you carried out any migration between CMSs?
- Can you evaluate the proposed methodology and its contribution?

The methodology was finally defined using a form similar to the current web standards and specifications. IETF recommendation RFC2119



Source: own processing

Figure 2: The application prototype part 1.



Source: own processing

Figure 3: The application prototype part 2.

as defined by Bradner (1997) is used in the text to signify the requirements. The words MUST, MUST NOT, SHOULD, SHOULD NOT, and MAY are used in the same manner.

Results and discussion

The methodology was updated according to the results of the latest research. The Creation and update, and Storage model modules were modified to comply with the needs for semantics. Additional modifications have been made according to the findings obtained during the Focus Group session and interviews.

For example, during the Focus Group session, users mentioned a need for offline mode for editing. The structuring of the content can be difficult at first sight. On the other side, use of templates can make it much more comfortable. To many users, adding the metadata descriptions was unclear and they did not see much sense in it. The following discussion lead to the recommendation, that the metadata enrichment should be automated as much as possible. The final application should also guide the author when adding it.

During the interviews, professionals and developers suggested, that there a broader discussion across involved companies and open source community should follow. There can be some requirements regarding the attributes in the storage model. Additionally, the professionals mentioned the problematic metadata section which requires users' activity. Another suggestion was for a future development. The model can be transformed to be more compatible with most used databases, such as MySQL, MongoDB, or Elasticsearch.

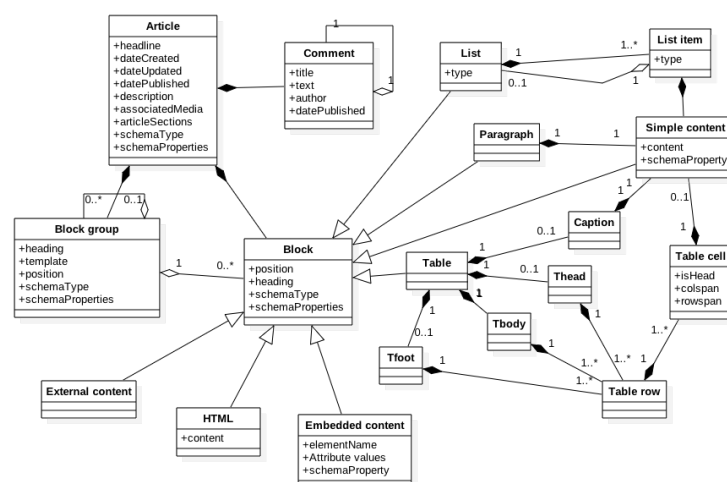
Formulation of the methodology

Creation and updating of the content

1. The content MUST be composed of independent blocks.
2. The blocks SHOULD NOT have any fixed position. Users SHOULD be able to easily change order of the blocks.
3. The preview of the result MUST be separated from the editor.
4. The preview SHOULD be shown to users in real-time.
5. The content MUST consist of several basic blocks which MUST respect the Storage model. The blocks are according to classes: Paragraph, List, Table, Embedded content, HTML (pure HTML).
6. The application SHOULD be able to create groups from abovementioned blocks to support more complicated content.
7. The content MAY be enriched by connecting external media, which can be specific for the implementing application.
8. The final application SHOULD provide an interface to define templates of the content.
9. The templates MAY set fixed position for blocks or groups (and create exception from the rule 2)
10. The templates and the editor SHOULD lead authors to properly add metadata descriptions

Content storage

1. The content MUST be stored according to the UML Domain model as shown in Figure 4.



Source: own processing

Figure 4: Class diagram domain model of the storage model.

Description of the classes is mentioned in Figure 4.

Presentation, semantics, share and archiving

1. Each item of the Block class **MUST** be enclosed in its appropriate HTML5 element. Blocks HTML and External content **MAY** be enclosed in div element.
2. Each block of the Block group class **MUST** consist of Block or Block group items. Composition is defined by the template.
3. Each instance of the Article class **MUST** be composed from Block and/or Block group items. Order is defined by the position attribute. The content is enclosed in Article element.
4. The structure inside article **MUST** respect semantic rules.
5. The content **MUST** respect the following structure at its minimum:
 - a. <header>
 - i. <h1>
 - ii. <time>
 - iii. <p>
 - b. <section>
 - c. <footer>
4. The final HTML result **SHOULD** be added by metadata description using one of the RDFa, JSON-LD, or Microdata format in connection with the most used vocabularies (i.e. Schema.org).
5. The defined templates **MUST** contain links to Blocks or Block groups using braces syntax – {Block_name}.
6. The template **MAY** contain HTML elements.
7. The implementing application **MUST** provide an exchange format in JSON, using the Schema defined <http://kit.pef.czu.cz/JSONschemas/wicm.json>
8. The application **SHOULD** use the format defined in the rule 9 to data exchange between server and client side.
9. Future versions of methodology and applications **MUST** be backward compatible with any older versions.

Conclusion

The proposed methodology has been divided into three main modules. The first module deals with content creation and updating. It consists of several recommendations and requirements

for the future applications, implementing the methodology. WYSIWYG editors are likely to be supplanted in the near future. To ease the authoring of the information content as well as enable its easier structuring, creation by composition from separate blocks presents itself as the logical next step. The core of the proposed methodology is content composition. The latest research has added the needs for semantics. The storage model has been enriched by necessary metadata description fields.

The Presentation, semantics, share and archiving module has been formulated. The need for use of semantics has been clarified. The necessary semantic HTML template structure has been defined. The module consists of a set of recommendations and requirements. Additionally, the exchange format for data export, transfer, share, and archiving has been defined. It uses JSON data format, complemented with JSON Schema definition.

The proposed methodology has many contributions. The structured composition of content can help authors produce superior and consistent content. The content can be more unified across the website as the environment would lead the user to do so. Additionally, the structuring natively leads the authors to produce responsive content which can be easily transformed for usage within applications for mobile devices.

The standardisation can help to develop general purpose libraries. The libraries can speed up further development of applications and content management systems. The export format can help to transfer the content between applications, upgrade, or in a transition to new application system. The use of general libraries and open source software is highly exploited within the areas of agriculture, regional and rural development. Therefore, the methodology contribution in this area is significant

Final consumers (readers of the content) can profit thanks to superior presentation of the content. The content can be also easily accessible through search engines and content aggregators thanks to better semantics.

Finally, all the mentioned contributions can lead to cost reduction for companies and organisations. Faster development saves the cost of programmers. Easier and more effective content creation and update can avoid hiring additional employees. More accessible content can bring more customers and income.

The methodology has been named as Web Information Content Management (WICM). For the future, the methodology will need to engage interest groups such as browser makers, content management systems producers, and other companies and organisations. A broader discussion and consensus are necessary to spread and finalize the methodology.

Acknowledgments

The results and knowledge included herein have been obtained owing to support from the following institutional grants. Internal grant agency of the Faculty of Economics and Management, Czech University of Life Sciences in Prague, grant no. 20181014 - "Permanent and globally unique identification of author entities".

Corresponding authors

Ing. Jan Masner, Ph.D.

Department of Information Technologies, Faculty of Economics and Management

Czech University of Life Sciences Prague, Kamýcká 129, 165 00 Praha – Suchbát, Czech Republic

E-mail: masner@pef.czu.cz

Orcid No. 0000-0003-4593-2306

References

- [1] Aghaei, S. (2012) "Evolution of the World Wide Web: From Web 1.0 to Web 4.0", *International Journal of Web & Semantic Technology*, Vol. January 2012. DOI 10.5121/ijwest.2012.3101.
- [2] Bally, L., Brittan, J. and Wagner, K. H. (1977) "A prototype approach to information system design and development", *Information & Management*, Vol. 1, No. 1, pp. 21-26. ISSN 0378-7206. DOI 10.1016/0378-7206(77)90005-2.
- [3] Bradner, S. (1997) "Key words for use in RFCs to Indicate Requirement Levels". Best Current Practice. [Online]. Available: <https://tools.ietf.org/html/rfc2119> [Accessed: 1 January 2017].
- [4] Brown, M. (2014) "Writing and Editing in the Browser", *The Journal of Electronic Publishing*, Vol. 17, No. 1. ISSN 1080-2711. DOI 10.3998/3336451.0017.111.
- [5] Choudhury, N. (2014) "World Wide Web and Its Journey from Web 1.0 to Web 4.0", *International Journal of Computer Science and Information Technologies*, Vol. 5, No. 6, pp. 8096-8100. ISSN 0975-9646.
- [6] Czerniak, B. (2015) "Paragraphs are Drupal's answer for structured content". [Online]. Available: <https://www.commercialprogression.com/post/paragraphs-are-drupals-answer-structured-content>. [Accessed: 20 January 2017].
- [7] Das, S., Goetz, M., Girard, L. and Clark, T. (2009) "Scientific publications on web 3.0", in *ELPUB 2009 - Rethinking Electronic Publishing: Innovation in Communication Paradigms and Technologies - Proceedings of the 13th International Conference on Electronic Publishing*, pp. 107-129. [Online]. Available: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84864867845&partnerID=40&md5=aaf22a5286a9eca030c3f6956c29f25b>. [Accessed: 15 March 2017].
- [8] Jorstad, I., Bakken, E. and Johansen, T. A. (2008) "Performance evaluation of JSON and XML for data exchange in mobile services", in *Winsys 2008: Proceedings of the International Conference on Wireless Information Networks and Systems*, pp. 237-240.
- [9] Khalili, A. and Auer, S. (2015) "WYSIWYM - Integrated visualization, exploration and authoring of semantically enriched un-structured content", *Semantic Web*, Vol. 6, No. 3, pp. 259-275. E-ISSN 2210-4968, ISSN 1570-0844. DOI 10.3233/SW-140157.
- [10] Khalili, A., Auer, S. and Hladky, D. (2012) "The RDFa content editor - From WYSIWYG to WYSIWYM", in *2012 IEEE 36th Annual Computer Software and Applications Conference*, pp. 531-540. DOI 10.1109/COMPSAC.2012.72.
- [11] Lin, B., Chen, Y., Chen, X. and Yu, Y. (2012) "Comparison between JSON and XML in Applications Based on AJAX", in *Proceedings - 2012 International Conference on Computer Science and Service System, CSSS 2012*, pp. 1174-1177. DOI 10.1109/CSSS.2012.297.

- [12] Masner, J., Jarolímek, J. and Kánská, E. (2018) "Novel Approach for Creation, Storage and Presentation of Online Information Content", *AGRIS on-line Papers in Economics and Informatics*, Vol. 10, No. 3, pp. 69-77. ISSN 1804-1930. DOI 10.7160/aol.2018.100306.
- [13] Merilinn, J. and Matinlassi, M. (2006) "State of the Art and Practice of OpenSource Component Integration", in 32nd *EUROMICRO Conference on Software Engineering and Advanced Applications (EUROMICRO'06)*, pp. 170-177. ISBN 0-7695-2594-6. DOI 10.1109/EUROMICRO.2006.61.
- [14] Minin, H. C., Alemán, J. J., Sacramento, C. and Trevisan, D. G. (2015) "A WYSIWYG editor to support accessible web content production", In Antona M., Stephanidis C. (eds) *Universal Access in Human-Computer Interaction. Access to Today's Technologies. UAHCI 2015. Lecture Notes in Computer Science*, vol 9175. Springer, Cham pp. 221-230. E-ISBN 978-3-319-20678-3, ISBN 978-3-319-20677-6. DOI 10.1007/978-3-319-20678-3_22.
- [15] Nath, K. and Iswary, R. (2016) "What Comes after Web 3.0 ? Web 4.0 and the Future", *International Conference on Computing and Communication Systems (I3CS'15)*, Vol. 2.
- [16] Nikolic, S. and Šilc, J. (2016) "Drupal 8 modules: Translation management tool and paragraphs", *Informatica (Slovenia)*, Vol. 40, No. 1, pp. 145-152.
- [17] Nurseitov, N., Paulson, M., Randall Reynolds, R. and Izurieta, C. (2009) "Comparison of JSON and XML Data Interchange Formats: A Case Study", *Scenario*, 59715, pp. 1-3. [Online]. Available: <http://www.cs.montana.edu/izurieta/pubs/caine2009.pdf>. [Accessed: 10 April 2017].
- [18] Peinl, R. (2016) "Semantic Web: State of the Art and Adoption in Corporations", *KI - Künstliche Intelligenz*. Vol. 30, No. 2, pp. 131-138, E-ISSN 1610-1987, ISSN 0933-1875. DOI 10.1007/s13218-016-0425-0.
- [19] Rudman, R. and Bruwe, R. (2016) "Defining Web 3.0: opportunities and challenges", *The Electronic Library*, Vol. 34, No. 1, pp. 116-131. ISSN 0264-0473. DOI 10.1108/EL-08-2014-0140.
- [20] Sahay, S. (2003) "Global software alliances : the challenge of 'standardization'", *Scandinavian Journal of Information Systems*, Vol. 15, No. 1, pp. 3-19. [Online]. Available: <http://aisel.aisnet.org/sjis/vol15/iss1/11>. [Accessed: 10 April 2017]. E-ISSN 1901-0990.
- [21] Sandrih, B., Tosić, D. and Filipović, V. (2017) "Towards Efficient and Unified XML/JSON Conversion - A New Conversion", *IPSI BGD Transactions on Internet Research*. IPSI Belgrade Ltd, Vol. 13, No. 1, Special Issue.
- [22] Stočes, M., Šilerová, E., Vaněk, J., Jarolímek, J. and Šimek, P. (2018) "Možnosti využití otevřených dat v sektoru cukr – cukrová řepa". *Listy cukrovarnické a řepářské*, Vol. 134, No. 3, pp. 117-121. ISSN 1210-3306 (in Czech).
- [23] Šimek, P., Stočes, M. and Vaněk, J. (2014) "Mobile access to information in the agrarian sector", *AGRIS On-line Papers in Economics and Informatics*, Vol.6, No. 2, pp. 89-96. ISSN 1804-1930.
- [24] Šimek, P., Vaněk, J., Stočes, M., Jarolímek, J. and Pavlík, J. (2017) "Mobile accessibility expense analysis of the agrarian WWW portal", *Agricultural Economics (Zemědělská ekonomika)*, Vol. 63, No. 5, pp. 197-203. ISSN 0139-570X (in Czech).
- [25] Srirama, S. N. (2017) "Mobile web and cloud services enabling Internet of Things", *CSI Transactions on ICT*. Vol. 5, No. 1, pp. 109-117. E-ISSN 2277-9086, ISSN 2277-9078. DOI 10.1007/s40012-016-0139-3.
- [26] Sweney, P. (2016) "The History of the Semantic Web is the Future of Intelligent Assistants", Medium.com. [Online]. Available: <https://medium.com/inventing-intelligent-machines/the-history-of-the-semantic-web-is-the-future-of-intelligent-assistants-da2ed50443be>. [Accessed: 18 March 2017].