

Towards a Visual Annotation Tool for End-User Semantic Content Authoring

Torgeir Lebesbye and Ahmet Soylu

Department of Informatics, University of Oslo, Norway
{torgeirl, ahmets}@ifi.uio.no

Abstract. There is a great amount of data on the Web, but to a large extent it is not published as linked data that computers can consume. Visual annotation tools have a considerable potential to empower end users to contribute their data to the Semantic Web, and could prove to be a solution to get more data on the Web linked. To this end, numerous tools have been developed; however, there still remains challenges to be addressed. In this paper, we present and discuss a set of prominent requirements toward the realisation of a visual annotation tool for end-user semantic content authoring.

Keywords: Semantic Content Authoring, Direct Manipulation Interface, End-User Development

1 Introduction

When Berners-Lee invented the Web in 1989, his motivation was to allow people to share and link documents without the barriers of hardware, file systems or data formats [4]. It later evolved into the Social Web, referred to as Web 2.0, where anyone could be the producers of contents through blogs, wikis and social media, and it became easy for people to collaborate on the Web.

In later years, the introduction of semantic technologies has made it possible to describe the meaning of data in a language more consumable for computers: *words* written with markup languages like XML¹, *grammar structure* using RDF², and *logic* described in knowledge languages like OWL³ [7]. This means a contextual Web where data in documents are linked and can be mashed with data from other sources in a completely new way. It is called the Semantic Web, often referred to as Web 3.0.

So-called lowercase semantic technologies enabled linked data to be added to documents with in-content annotations [1]. Microformats⁴ is a widely used family of data formats that includes hCard, hCalendar and hAtom. eRDF was W3C's original attempt at simplifying annotation, but they abandoned it for

¹ <http://www.w3.org/TR/xml/>

² <http://www.w3.org/TR/rdf-schema/>

³ <http://www.w3.org/TR/owl-primer/>

⁴ microformats.org

RDFa⁵. Microdata⁶ is an alternative to RDFa originating from the WHATWG initiative⁷.

While linking documents and creating social content is easy and accessible through numerous tools and services, adding semantics to a web document requires knowledge of the involved technologies and a large technical skill-set that most end users often do not possess. Most of the data published on the Web isn't linked, and much of this data is managed by end users within organisations and on the open Web. Therefore, if we are to convert the current Web dominated by unstructured documents into a Web of Data, end-user involvement has a crucial role to play. In this respect, visual annotation tools have a considerable potential to empower end users to contribute their data to the Semantic Web, and could prove to be a solution to get more data on the Web linked.

To this end, numerous tools have been developed [11]; however, there still remain challenges to be addressed. In this paper, we present and discuss a set of prominent requirements towards the realisation of a visual annotation tool for end-user semantic content authoring. For each requirement, we shortly discuss how it improves a visual annotation tool from an end-user perspective. We believe that the discussion presented in this paper may be useful for researchers and practitioners working on annotation tools for semantic content authoring.

The rest of the paper is organised as follows: Section 2 discusses design requirements for a visual annotation tool for end users. Section 3 looks at related work, and finally Section 4 concludes the paper.

2 Design Requirements

End-User Development (EUD) allows users to act as non-professional software developers, creating, modifying, or extending software artefacts [14]. It includes spreadsheets and filters for emails, and is something more and more users do without thinking of it as software development. Not surprisingly, studies has shown the number of end-user programmers vastly outnumber professional programmers [19] and estimate it will continue to do so in the future. Annotation tools for end users are meant to enable end users to modify and extend software artefacts, and could be considered within EUD.

A visual annotation tool employs a direct manipulation approach [20], where end users can directly manipulate visual objects representing domain elements and application functionality to incorporate semantic knowledge, rather than dealing with a command language. We can assume that users managing web pages have at least some domain knowledge. An annotation tool should empower them to access and use this domain knowledge without requiring expertise in web technologies and ontologies, and should take the following into account:

⁵ <http://www.w3.org/TR/rdfa-syntax/>

⁶ <http://www.w3.org/TR/microdata/>

⁷ <https://whatwg.org/>

2.1 Bottom-up

Semantic content authoring tools are often divided into two main categories: top-down and bottom-up [11]. A top-down approach focuses on making and extending ontologies during the annotation process, while a bottom-up approach focuses on the document and uses existing ontologies to annotate the document.

End users are expected to have little or no experience with ontologies, therefore a bottom-up approach is preferable. Moreover, lifting unstructured content to a semantic level is an important issue, given that today the Web is dominated by unstructured documents.

2.2 Human-driven

Another issue is the level of automation – machine-driven vs. human-driven approaches. Some tools detect and annotate text automatically largely based on natural language processing (NLP) techniques [13]. Others provide suggestions, and keep the document valid during the editing process. Machine-driven approaches have a huge advantage in annotation speed, while human-driven approaches hold a higher annotation quality. End users might have little experience with semantic technologies, but as managers of the content they more often have some domain-knowledge. A full automation does not take advantage of this knowledge.

Automation support is particularly important when a large number of documents are involved, yet this should be adequately intertwined with a manual approach. Considering typical users and documents on the Web, a human-driven approach remains more accessible as a generic solution, since automated approaches usually require domain-specific configurations.

2.3 Exploration support

An annotation tool for end users should feature an exploration support, where the underlying ontologies and data can be explored visually. In this respect, visualisations are a powerful way to make the content of a service or tool accessible. For end users, visualisations can make it easier to understand the underlying domain in terms of concepts, properties, and instances.

The big challenge of ontology visualisation is scalability. Ontologies vary in size, from a few hundred nodes to hundreds of thousands of nodes. The vast size of a large ontology can be intimidating, and very hard to get an overview of. A survey on ontology visualisation methods [10] suggests that visualisations should be coupled with effective search functionality and take advantage of semantic data and user data to make the ontology exploration more efficient.

2.4 Complete editing suite

Some annotation tools available today remain primitive, lack important functionality and expressivity. Many of the fully automated tools do not preserve

change in a document. Some tools do not support different in-content annotation technologies and most do not support linking entities through objects' properties. Having said that, a visual annotation tool is not expected to be fully expressive, as certain functionalities and ontology constructs are difficult to grasp even in a visual form.

An annotation tool should preserve change, support the most important formats, and allow for full editing of the document data, while adequately managing the trade-off between usability and expressivity. To this end, often and commonly used functionality and ontology constructs have to be identified and classified with respect to their complexity, as perceived by the end users.

2.5 Usability evaluation

Most of the tools developed in academia undergo little end-usability evaluation, and user studies is often limited to students affiliated with the research groups. Qualitative end-user studies measure whether a tool is competent of meeting its identified aim with respect to a set of criteria, such as effectiveness (i.e., completeness and accuracy) and efficiency (i.e., the cost associated such as time), user satisfaction, learnability etc. [5].

A successful validation of usability and functionality requires user studies to be conducted with users that match the profile of target end users. However, one also needs to be aware that a single summative study only at the end is not sufficient; several formative end-user studies should be held with intermediary prototypes during the design and development process for timely identification of any usability problems.

3 Related Work

Existing tools vary in approach, level of automation, and domain-dependency, in what follows we address only the most prominent ones. Most tools extend the software that the targeted users already uses, to piggyback on concepts and workflow that the users already are familiar with, such as widely used WYSIWYG editor TinyMCE⁸ and wiki tool MediaWiki⁹.

HayStack semantic blogging [9] and semiBlog [16] were early works on enabling end users to add metadata about the structure and content of their blog posts, their relations (in reply), and subscription to blogs using RSS. DataPress [3] and LinkedBlog [18] are both extensions of the WordPress blogging tool, and the annotation tools are integrated as WYSIWYG editors.

SweetWiki [6] and Semantic MediaWiki [8] are semantic content authoring tools built as wikis that follow the bottom-up approach. They are both based on MediaWiki, a wiki tool originating from the Wikipedia project. They are made specifically for the wiki domain, both with the goal of enhancing the content in the wiki with semantics, while improving search results and other core functionalities.

⁸ www.tinymce.com/

⁹ <https://www.mediawiki.org/>

Loomp [15] is an annotation tool for journalists, where the user annotates information fragments, makes mash-ups of those fragments, and keeps track of and reuses them for semantic linking and search.

Annotation tools for Content Management System (CMS) tend to be fully automated. Epiphany [2] uses a web service that finds instances in a web page, and automatically returns a version of the web page with RDFa annotations. It uses the light-box effect often used for image galleries to visualise embedded RDFa. FLERSA [17] is an automated annotation tool built upon Joomla, a popular CMS for building web portals.

OntosFeeder [13] is a web service made to be integrated with a CMS. It supports TinyMCE and FCKeditor on both WordPress and Drupal, is independent of both the editor and the CMS, and is fully automated.

RDFaCE[12] is another extension of the TinyMCE. It has four synchronised views (WYSIWYG editing view, annotations view, fact view, and HTML/RDFa source view), and allows users to switch freely between them during the editing process. It does, however, not include an exploration support of the underlying ontology and does not support linking entities through object properties.

4 Conclusion

In this paper, we have looked at design requirements for a visual annotation tool for end-user semantic content authoring. We believe that such a tool should follow a bottom-up approach, be human-driven, use visualisations to facilitate document annotation and the exploration of the underlying ontologies, and undergo user studies with a representative set of end users. The related work suggests that existing approaches mostly fail to meet these requirements.

Our future work involves design and development of a visual annotation tool for end-user semantic content authoring with these principles in mind.

References

1. B. Adida. hGRDDL: Bridging microformats and RDFa. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(1):54–60, 2008.
2. B. Adrian, J. Hees, I. Herman, M. Sintek, and A. Dengel. Epiphany: Adaptable rdfa generation linking the web of documents to the web of data. In *Proceedings of the 17th International Conference on Knowledge Engineering: Practice and Patterns (EKAW 2010)*, volume 6317 of *LNAI*, pages 178–192. Springer, 2010.
3. E. Benson, A. Marcus, F. Howahl, and D. Karger. Talking about data: Sharing richly structured information through blogs and wikis. In *Proceedings of the 9th International Semantic Web Conference (ISWC 2010)*, volume 6496 of *LNCS*, pages 48–63. Springer, 2010.
4. T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web - a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, 284(5):34–43, 2001.
5. N. Bevan and M. Macleod. Usability measurement in context. *Behaviour and Information Technology*, 13(1-2):132–145, 1994.

6. M. Buffa, F. Gandon, G. Ereteo, P. Sander, and C. Faron. SweetWiki: A semantic wiki. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(1):84–97, 2008.
7. B. C. Grau, I. Horrocks, B. Motik, B. Parsia, P. Patel-Schneider, and U. Sattler. OWL 2: The Next Step for OWL. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4):309–322, 2008.
8. D. M. Herzig and B. Ell. Semantic mediawiki in operation: Experiences with building a semantic portal. In *Proceedings of the 9th International Semantic Web Conference (ISWC 2010)*, volume 6497 of *LNCS*, pages 114–128. Springer, 2010.
9. D. R. Karger and D. Quan. What would it mean to blog on the semantic web? In *Proceedings of the 3rd International Semantic Web Conference (ISWC 2004)*, volume 3298 of *LNCS*, pages 214–228. Springer, 2004.
10. A. Katifori, C. Halatsis, G. Lepouras, C. Vassilakis, and E. Giannopoulou. Ontology visualization methods—a survey. *ACM Computing Surveys*, 39(4):10, 2007.
11. A. Khalili and S. Auer. User interfaces for semantic authoring of textual content: A systematic literature review. *Web Semantics: Science, Services and Agents on the World Wide Web*, 22:1–18, 2013.
12. A. Khalili, S. Auer, and D. Hladky. The RDFa Content Editor - From WYSIWYG to WYSIWYM. In *Proceedings of the IEEE 36th Annual Computer Software and Applications Conference (COMPSAC 2012)*, pages 531–540. IEEE, 2012.
13. A. Klebeck, S. Hellmann, C. Ehrlich, and S. Auer. Ontosfeeder—a versatile semantic context provider for web content authoring. In *Proceedings of the 8th Extended Semantic Web Conference (ESWC 2011)*, volume 6644 of *LNCS*, pages 456–460. Springer, 2011.
14. H. Lieberman, F. Paternó, M. Klann, and V. Wulf. End-User Development: An Emerging Paradigm. In H. Lieberman, F. Paternó, and V. Wulf, editors, *End-User Development*, volume 9 of *Human-Computer Interaction Series*, pages 1–8. Springer, Netherlands, 2006.
15. M. Luczak-Rösch and R. Heese. Linked Data Authoring for Non-Experts. In *Proceedings of the WWW2009 Workshop on Linked Data on the Web (LDOW 2009)*, volume 538 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2009.
16. K. Möller, U. Bojars, and J. G. Breslin. Using semantics to enhance the blogging experience. In *Proceedings of the 3rd European Semantic Web Conference on the Semantic Web: Research and Applications (ESWC 2006)*, volume 4011 of *LNCS*, pages 679–696. Springer, 2006.
17. J. L. Navarro-Galindo and J. Samos. The FLERSA tool: adding semantics to a web content management system. *International Journal of Web Information Systems*, 8(1):73–126, 2012.
18. I. Ruiz-Rube, C. M. Cornejo, J. M. Doderó, and V. M. García. Development issues on linked data weblog enrichment. In *Proceedings of the 4th International Conference on Metadata and Semantic Research (MTSR 2010)*, volume 108 of *CCIS*, pages 235–246. Springer, 2010.
19. C. Scaffidi, M. Shaw, and B. Myers. Estimating the numbers of end users and end user programmers. In *Proceedings of the 2005 IEEE Symposium on Visual Languages and Human-Centric Computing (VLHCC 2005)*, pages 207–214. IEEE, 2005.
20. B. Shneiderman. Direct Manipulation: A Step Beyond Programming Languages. *Computer*, 16(8):57–69, 1983.