RDFaCE-Lite: A WYSIWYM Editor for User-Friendly Semantic Text Authoring

Ali Khalili(✉) and Sören Auer

IFI/BIS/AKSW, Universität Leipzig, Johannisgasse 26, 04103 Leipzig, Germany
{khalili,auer}@informatik.uni-leipzig.de
http://aksw.org

Recently practical approaches for managing and supporting the life-cycle of semantic content on the Web of Data made quite some progress. However, the currently least developed aspect of the semantic content life-cycle is the user-friendly manual and semi-automatic creation of rich semantic content. In this demo we will present the RDFaCE-Lite editor and will show:

- how users can annotate textual content using vocabularies and named entities published on the Data Web
- how different NLP APIs can be combined in order to maximize precision and recall of the annotation process.
- how the RDFaCE-lite annotation environment can be used within existing applications such as Blogs, CMSs, etc.

RDFaCE-Lite combines WYSIWYG text authoring with the creation of rich semantic annotations. WYSIWYG text authoring is meanwhile ubiquitous on the Web and part of most content creation and management workflows. It is part of Content Management Systems, Weblogs, Wikis, fora, product data management systems and online shops, just to mention a few. Our goal with this work is to integrate the semantic annotation directly into the content creation process and to make the annotation as easy and non-intrusive as possible. The RDFaCE-Lite implementation is open-source and available for download together with an online demo at http://aksw.org/Projects/RDFaCE.

RDFaCE-Lite is developed as a plug-in for *TinyMCE Rich Text Editor* (http://tinymce.moxiecode.com). This open source HTML editor was chosen because it is very flexible to extend and is used in many popular Content Management Systems (CMS), blogs, wikis and discussion forums, etc. Therefore, by focusing efforts on this one particular editor, it is possible to quickly propagate accessible semantic content authoring practices to a number of other tools. As depicted in Fig. 1, RDFaCE-Lite provides one click text annotation by employing the following components:

NLP APIs Abstraction and Integration. Starting to annotate a document from scratch is very tedious and time consuming. There are already some Natural Language Processing (NLP) APIs available on the Web which extract specific entities and relations from the text. By using these APIs, we can provide a good starting point for further user annotations. Users then can modify and extend this automatically pre-annotated content. RDFaCE-Lite currently uses

E. Simperl et al. (Eds.): ESWC 2012 Satellite Events, LNCS 7540, pp. 376–380, 2015.
DOI: 10.1007/978-3-662-46641-4_30

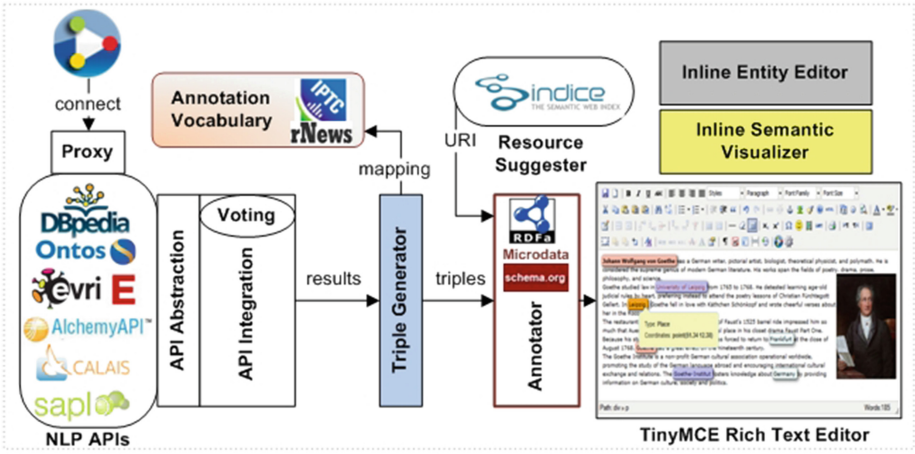

Fig. 1. RDFaCE-Lite system architecture.

the *OpenCalais*, *Ontos*, *Alchemy*, *Extractiv*, *Evri*, *Lupedia*, *DBpedia Spotlight* and *Saplo*[1] APIs to enrich the text. Since each of these APIs use a different connecting interface as well as a different data structure for output we have implemented a *Proxy* and *API Abstraction* component. Proxy performs the connecting task by providing a separate adapter for each API. Abstraction component unifies the output of each API to a standard format[2] used by RDFaCE-Lite.

Besides annotation by each individual API, RDFaCE-Lite supports combining the results of multiple NLP APIs which yields superior performance compared to each individual (cf. http://rdface.aksw.org/samples/results.html). For this purpose, we have implemented an *API Integration* component which uses voting algorithm to integrate the results of different NLP APIs. This feature is fully configurable. Users can select their desired NLP APIs plus the number of agreements in their setting preferences.

Triple Generator. This component is responsible for generating the RDF triples to be embedded in the text. To achieve this goal, triple generator employs different existing vocabularies. In RDFaCE-Lite we use rNews 1.0 (http://dev.iptc.org/rNews) vocabulary as our annotation schema. *rNews* is a proposed standard to annotate HTML documents with news-specific metadata. rNews is proposed by International Press Telecommunications Council (IPTC) which is a consortium of the world's major news agencies, publishers and industry vendors. All the entities and properties extracted by NLP APIs are mapped to their corresponding ones in the rNews vocabulary.

[1] OpenCalais - http://www.opencalais.com, Ontos - http://www.ontos.com, Alchemy - http://www.alchemyapi.com, Extractiv - http://extractiv.com, Evri - http://www.evri.com, Lupedia - http://lupedia.ontotext.com/ and DBpedia Spotlight - http://dbpedia.org/spotlight.

[2] NLP Interchange Format (NIF) available at http://nlp2rdf.org/.

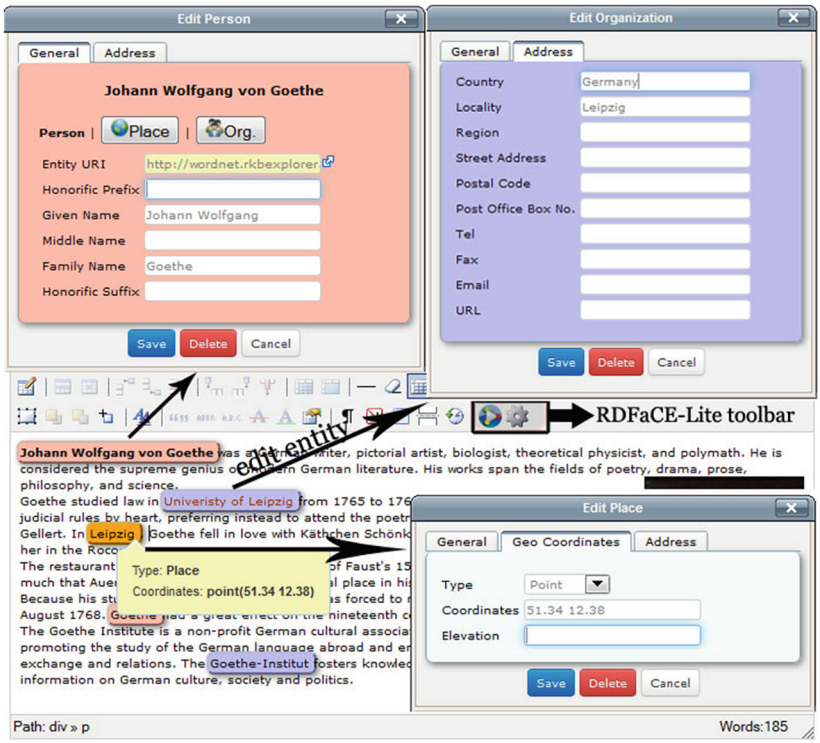

Fig. 2. Editing entity annotations in RDFaCE-Lite.

Annotator. This component manipulates the Document Object Model (DOM) according to the generated triples. Annotator uses the resource suggester component to generate URIs for entities. Sindice (http://sindice.com) semantic search engine is employed by resource suggester. The annotator component supports *RDFa 1.1* and *Microdata* (based on Schema.org) annotation formats.

Inline Semantic Visualizer. This component provides a WYSIWYM (What-You-See-Is-What-You-Mean) view on top of the WYSIWYG (What-You-See-Is-What-You-Get) view. The WYSIWYG view is the classical interface for rich-text authoring and used by authors, journalists etc. WYSIWYG text authoring is meanwhile ubiquitous on the Web and part of most content creation and management workflows. Users authoring content are used to interact with a WYSIWYG views and there exists a wide variety of WYSIWYG editors and editing components, which can be used on the Web or offline.

The WYSIWYM view is an extension of the WYSIWYG view, which highlights named entities and other semantic information. The highlighting is realized with special CSS3 selectors for the RDFa annotations. They are thus easily configurable in terms of color borders, backgrounds etc. When pointing with the mouse on a highlighted annotation RDFaCE shows additional information concerning

the particular annotation as a dynamic tooltip. RDFaCE also supports editing in the WYSIWYM view by letting a user select entities he wants to annotate and provisioning of respective annotation functionality either via the context menu or a specific form, which opens as an overlay.

Inline Entity Editor. Editing entity annotations by one click is the main task of this component. Figure 2 shows the inline editor window for Place, Person and Organization entities. Inline editor creates forms for each class of the recognized entity properties. For instance, Place entity has three categories of properties namely General, Geo-Coordinates and Address. For each of them users can fill in the values in the corresponding form.

The RDFaCE-Lite tool is very versatile and can be applied in a vast number of use cases. *Data-driven Journalism and Semantic Blogging* are two main use cases of RDFaCE-Lite. Data-driven journalism and semantic blogging deal with open data that is freely available online and analyzed with open source tools. Semantically annotated news and blog posts provided by RDFaCE-Lite facilitate a number of important aspects of information management:

- For *search and retrieval* enriching documents with semantic representations helps to create more efficient and effective search interfaces, such as faceted search or question answering.
- In *information presentation* semantically enriched documents can be used to create more sophisticated ways of flexibly visualizing information, such as by means of semantic overlays.
- For *information integration* semantically enriched documents can be used to provide unified views on heterogeneous data stored in different applications by creating composite applications such as semantic mashups.
- To realize *personalization*, semantic documents provide customized and context-specific information which better fits user needs and will result in delivering customized applications such as personalized semantic portals.
- For *reusability* and *interoperability* enriching documents with semantic representations facilitates exchanging content between disparate systems.

RDFace-Lite is published as a plug-in for WordPress blogging platform[3] thereby facilitates the semantic blogging process. Wordpress is often customized into a Content Management System (CMS) and is used by over 14 % of the 1,000,000 biggest websites (54.4 % of CMS market share) [7] which would potentially advance the promotion of semantic content even more. Furthermore, RDFaCE-Lite supports rNews standard which is getting adopted by the popular news providers such as NYTimes. Using this news-specific vocabulary to annotate entities and relationships between them will facilitate data journalism.

Regarding the related work, there are already many tools available for semantic text authoring. *WYMeditor*[4], *DataPress* [1], *Loomp* [3], *FLERSA* [4], *RDFauthor* [6] and *SAHA* 3 [2] are some examples of available tools. None of these tools

[3] Available at http://wordpress.org/extend/plugins/rdface/.
[4] http://www.wymeditor.org.

support Microdata annotation format Among the tools, RDFauthor and Loomp are adopting a similar approach as RDFaCE-Lite but do not provide any feature for automatic content annotation.

The RDFauthor approach is based on the idea of making arbitrary XHTML views with integrated RDFa annotations editable [6]. RDFauthor converts an RDFa-annotated view directly into an editable form thereby hiding the RDF and related ontology data models from novice users. The main difference between RDFaCE-Lite and RDFauthor is that RDFauthor assumes that the RDFa content is already existing while RDFaCE provides a complementary feature to create new RDFa annotations.

Loomp is another related tool representing a proof-of-concept for the *One Click Annotation* (OCA) strategy. The Web-based OCA editor allows for annotating words and phrases with references to ontology concepts and for creating relationships between annotated phrases. The main difference between Loomp and RDFaCE is that Loomp relies on the functionality of a server managing the semantic content while RDFaCE-Lite provides client-side annotation for modifying semantic content directly.

NERD [5] as a partially related work is an evaluation framework which records and analyzes ratings of Named Entity extraction and disambiguation tools. The main difference between RDFaCE-Lite and NERD is that RDFaCE-Lite employs the voting approach to combine the results of NLP APIs for automatic annotation but NERD expects a human being to manually compare the results of different NLP APIs and choose the right one for annotation. Furthermore, NERD does not focus on the annotation and authoring task but more on evaluating NLP APIs.

References

1. Benson, E., Marcus, A., Howahl, F., Karger, D.: Talking about data: sharing richly structured information through blogs and Wikis. In: Patel-Schneider, P.F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J.Z., Horrocks, I., Glimm, B. (eds.) ISWC 2010, Part I. LNCS, vol. 6496, pp. 48–63. Springer, Heidelberg (2010)
2. Frosterus, M., Hyvönen, E., Laitio, J.: DataFinland—a semantic portal for open and linked datasets. In: Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., De Leenheer, P., Pan, J. (eds.) ESWC 2011, Part II. LNCS, vol. 6644, pp. 243–254. Springer, Heidelberg (2011)
3. Luczak-Roescht, R.H.M.: Linked data authoring for non-experts. In: Proceedings of the WWW 2009, Workshop Linked Data on the Web (2009)
4. Navarro-Galindo, J.L., Samos, J.: Manual and automatic semantic annotation of web documents: the FLERSA tool. In: iiWAS 2010, pp. 542–549. ACM, New York (2010)
5. Rizzo, G., Troncy, R.: NERD: a framework for evaluating named entity recognition tools in the web of data (2011)
6. Tramp, S., Heino, N., Auer, S., Frischmuth, P.: RDFauthor: employing rdfa for collaborative knowledge engineering. In: Cimiano, P., Pinto, H.S. (eds.) EKAW 2010. LNCS, vol. 6317, pp. 90–104. Springer, Heidelberg (2010)
7. W3Techs. Usage of content management systems for websites, June 2011