Applications of Semantic Publishing

Borislav Popov
Ontotext AD
`borislav.popov@ontotext.com`

1   Abstract

In recent years Semantic publishing applications get more and more domain-spread and user-oriented in several aspects, among which: customization and re-purpose of data and content reflecting the user needs in various domains; focused summaries with respect to user interests; high relevance of the retrieved information and minimal effort in receiving it.

There are various works, exploring the relation between publishing and Linked Open Data, since the latter enriches the semantics successfully across various domains. In (Villazon-Terrazas et. al 2012), for example, authors present their idea on a life cycle model (specification, modeling, generation, linking, publication, exploitation) and demonstrate its application within various domains. At the same time, in (Mendes et. al 2011) a DBpedia service has been presented (called DBpedia Spotlight), which automatically annotates text documents with DBpedia URI's using the DBpedia in-house ontology. Similarly, Zemanta[1] provides a plug-in to content creators, which recommends links to relevant content (articles, keywords, tags). Our approach is generally in-line with these ideas and services – domain specific applications, automatic semantic annotation, and addition of relevant linked content. However, our focus is preferably on: the trade-off between the semantic knowledge holders (ontologies, linked data) and their language reflection (domain texts), mediated by the linguistic processing pipelines; the adaptive flexibility of the constructed applications and the efficient storage and publishing of large data.

Within Ontotext, examples of mass media, semantic publishing web sites, such as the BBC's sport web[2] and the official web of the London's Olympics 2013, have proven to attract a multi-million user bases. Behind such applications, as revealed by lead engineers at the BBC[3], there lies the complex architecture of the state-of-the-art Semantic and Text Analytics technologies, such as in-house: fast RDF database management system OWLIM[4] and knowledge management platforms KIM[5]; for robust semantic annotation and search, as well as for text analytics applications.

Both platforms are incorporated into numerous successful Semantic Publishing Solutions (including the BBC Sport[6], Press Association[7], Newz[8], EuroMoney[9], Publicis[10] etc.). For the core methodology see (Kiryakov et. al 2003) and (Popov et. al 2003). Starting with the FIFA 2010 BBC web site, through the London Olympics, feeding the official news site with enriched content together with Press Association, we have built domain expertise, sound solution implementation methodologies, and a semantic publishing platform to serve our clients. Beyond mass media, specialized publishers licensed our products and commissioned us for customizations - like Euromoney for macroeconomic report analytics and Oxford University Press and IET for high value scientific content.

This talk aims to describe the parameters of our domain adaptation approach, used successfully in many projects for more than 5 years, to build rigorous semantic publishing solutions.

Our strategy relies on the calibration between the RDF semantic repository OWLIM, the semantic resources in KIM and the optimized Text Analytics techniques including methodologies for fast creation of gold data in the selected do-

---

[1] http://en.wikipedia.org/wiki/Zemanta
[2] www.bbc.com/sport

[3] www.bbc.co.uk/blogs/bbcinternet/2012/04/sports_dynamic_semantic.html
[4] www.ontotext.com/owlim
[5] http://www.ontotext.com/kim
[6] http://www.ontotext.com/publishing
[7] http://www.pressassociation.com/
[8] newz.nl
[9] http://www.euromoney.com/
[10] http://www.publicis.de/

main; focused curation of the automatically analyzed data and the application of advanced machine learning algorithms in data clustering. Thus, the success of our solutions lays in the customization of the advanced semantic technologies in combination with text analytics techniques, tuned to the needs of publishers and adapted to the requested domains.

2 Short Bio

Borislav Popov is the Head of the Semantic Annotation and Search division at Ontotext. He has specialized in AI, spent some time on landmark projects in the financial and ERP industry across the Balkans with clients like BASF and AC Nielsen. He is a part of Ontotext since its founding and leads the company's involvement in several EC funded projects with multi–million budgets. He took part in the birth of the KIM Platform and since then is leading both its development, the semantic annotation and search division and is primarily responsible for all the solution the group provides. Under his guidance the group delivered multiple solutions in Publishing and Media for the BBC, Press Association and several other major customers.

References

Atanas Kiryakov, Borislav Popov, Damyan Ognyanoff, Dimitar Manov, Angel Kirilov, Miroslav Goranov. 2003. Semantic Annotation, Indexing, and Retrieval. In: *2nd International Semantic Web Conference (ISWC2003)*, 20-23 October 2003, Florida, USA. LNAI Vol. 2870, pp. 484-499, Springer-Verlag Berlin Heidelberg 2003.

Pablo N. Mendes, Max Jakob, Andres Garcia-Silva and Christian Bizer. 2011. DBpedia spotlight: shedding light on the web of documents. In: *Proceedings of the 7th International Conference on Semantic Systems*, pp. 1-8, ACM, New York, NY, USA.

Borislav Popov, Atanas Kiryakov, Angel Kirilov, Dimitar Manov, Damyan Ognyanoff and Miroslav Goranov. 2003. KIM – Semantic Annotation Platform. In: *2nd International Semantic Web Conference (ISWC2003)*, 20-23 October 2003, Florida, USA. LNAI Vol. 2870, pp. 834-849, Springer-Verlag Berlin Heidelberg 2003.

Boris Villazon-Terrazas, Daniel Vila-Suero, Daniel Garijo, Luis M. Vilches-Blazquez, Maria Poveda-Villalon, Jose Mora, Oscar Corcho, and Asuncion Gomez-Perez. 2012. Publishing Linked Data - There is no One-Size-Fits-All Formula. In: *Proceedings of the European Data Forum 2012, Copenague, Dinamarca*.