

The BBC's Dynamic Semantic Publishing Showcase

By using Ontotext technology, the BBC has achieved dynamic metadata-driven semantic publishing for their FIFA 2010 World Cup Web site, as described in Jem Rayfield's SemTech 2011 presentation.



BBC Future Media & Technology department have been transforming the BBC relational content management model and static publishing framework to a fully dynamic semantic publishing architecture. With minimal journalistic management, assets are being enriched with links to ontology concepts. This novel semantic approach provides an improved navigation, content re-use and re-

purposing, and supports advertisement based on semantic similarity. This allows automatic aggregation and rendering of links to relevant stories and improves the user experience in the site.

"A high-performance dynamic semantic publishing framework facilitates the publication of automated metadata-driven web pages that are light-touch, requiring minimal journalistic management, as they automatically aggregate and render links to relevant stories". -- Jem Rayfield, Senior Technical Architect

The challenge

The challenge is to create adequate and maximally useful web presence for publishers due to the volume, diversity and the dynamics of their content.

Dynamic online publishing uses content metadata to automatically generate and update web pages. Semantic metadata allows for deeper and more informative web presence. Content can be categorized, positioned, interlinked and navigated taking into account some of its semantics.

"The goal is to be able to more easily and accurately aggregate content, find it and share it across many sources. From these simple relationships and building blocks you can dynamically build up

incredibly rich sites and navigation on any platform." -- John O'Donovan, Chief Technical Architect, BBC Future Media & Technology

The solution: Dynamic Semantic Publishing Framework

The Dynamic Semantic Publishing (DSP) architecture of the BBC curates and publishes HTML and RDF aggregations based on embedded Linked Data identifiers, ontologies and associated inference. RDF semantics improve navigation, content re-use, re-purposing, search engine rankings, journalist determined levels of automation (“edited by exception”) and support semantic advertisement placement for audiences outside of the UK. The DSP approach facilitates multi-dimensional entry points and a richer navigation, greatly improving user experience and levels of engagement.



OWLIM serves as the RDF store in the architecture facilitating metadata-based management, inference and retrieval of content.

“A RDF triplestore and SPARQL approach was chosen over and above traditional relational database technologies due to the requirements for interpretation of metadata with

respect to an ontological domain model”. --Jem Rayfield, Senior Technical Architect, BBC News and Knowledge

First Real Proof: The BBC's World Cup 2010 Web Site

BBC's 2010 World Cup website was the first project running on its Dynamic Semantic Publishing architecture.



"The World Cup site is a large site with over 700 aggregation pages (called index pages) designed to lead you on to the thousands of story pages and content

...we are not publishing pages, but publishing content as assets which are then organized by the metadata dynamically into pages, but could be re-organized into any format we want much more easily than we could before."

... The index pages are published automatically. This process is what assures us of the highest quality output, but still save large amounts of time in managing the site and makes it possible for us to efficiently run so many pages for the World Cup."

John O'Donovan, Chief Technical Architect, BBC Future Media & Technology

Some statistics about BBC's World Cup web site follow below:

- 800+ Dynamic aggregations/pages (Player, Team, Group, etc.), generated through SPARQL queries
- Average unique page requests/day: 2 million
- Average SPARQL queries/day: 1 million
- 100s repository updates/inserts per minute with OWL 2 RL reasoning
- Multi data center fully resilient, clustered 6 node triple store

Further Developments

After the success of the World Cup 2010 project Ontotext was also awarded a contract to implement another important part of the DSP architecture of BBC - the Concept Extraction service.

In 2011 the DSP framework is being adopted in the SPORT section of the web site of the BBC. Future plans are to adopt the platform for the BBC Olympics 2012 online coverage.

Testimonials

"We are happy to have chosen BigOWLIM, after evaluation of several RDF stores, to build a high-performance semantic stack for the World Cup 2010 site"

Jem Rayfield, Senior Technical Architect

"... the first large scale, mass media site to be using concept extraction, RDF and a Triple store to deliver content".

John O'Donovan, Chief Technical Architect, BBC Future Media & Technology

"It's really fantastic to see organizations like the BBC building really exciting semantic applications... It Begins ...": comments at the Read Write Web's post on the subject

A talk about the BBC transformational technology strategy for evolving their relational content model and static publishing framework to a fully dynamic semantic publishing (DSP) architecture, specifically focusing on News, Sport & Knowledge products, will be given by Jem Rayfield at the 2011 Semantic Technology Conference, June 5-9, California.