

WordLift: Meaningful Navigation Systems and Content Recommendation for News Sites running WordPress

Andrea Volpini and David Riccitelli

WordLift (WL) is an extension of WordPress that helps to search, organise, tag, browse, evaluate, and share content produced by bloggers, content owners and independent news organisations.

While traditional CMS contain only text which computers can neither understand nor evaluate, WordLift adds semantic annotations that allow a website or a blog to function as a collaborative knowledge base. WordLift combines information publicly available as linked open data to support the editorial workflow of web publishers by suggesting relevant information, images and links.

WordLift analyses articles using Named Entity Recognition (NER). The entities may pertain to different vocabulary sets including but not limited to DBpedia, GeoNames and Freebase. WordLift also provides UIs for creating and curating custom vocabularies.

While annotating contents editors can identify the basic '*who, what, when and where*' of an article and structure information around it by creating new entities in their custom vocabularies.

Named entities are stored in the local WordPress database as well as in an optimized triple store in the cloud running Apache Marmotta. Annotation and entities are accessible via a Web Page and also using RDF, N3 and JSON-LD formats. Moreover the triple-store can be queried via SPARQL.

In the context of FP7-MICO¹ an EU co-funded R&D project aiming at analysing “**media in context**” by orchestrating different content extraction tools we’re extending the capabilities of WordLift to support news organisations in order to:

- reduce the time spent in creating news by suggesting relevant media and links to related content (activity otherwise done manually by the editor)
- increase page views per session by adding:
 - a faceted search UI for browsing content mentioning the same entity
 - a widget providing content based recommendation leveraging on the network of extracted entities (item based recommendation)
 - a widget providing content based recommendation based on the user profile (user based recommendation)
- increase and qualify the traffic by adding JSON-LD markup to articles, images and videos.

In the demonstration a real-world use case from Greenpeace Italy will be presented.

The environmental organisation is experimenting with WordLift to increase its website traffic, the loyalty of its readership and the user retention.

The plugin of WordLift currently under development is version 3. The service of WordLift version 2 is suspended as InSideOut10 is focusing all its efforts to bring version 3 live.

The source code of version 3 is available on GitHub:

- <https://github.com/insideout10/wordlift-plugin> (PHP)
- <https://github.com/insideout10/wordlift-plugin-js> (Javascript)
- <http://insideout10.github.io/wordlift-plugin/docs/namespaces/default.html> (developer documentation)

¹ MICO - Media In Context - <http://www.mico-project.eu>

About InSideOut10

InSideOut10 is an *Italian start-up* and consulting firm with an in-depth experience on web publishing and media delivery platforms. InSideOut10 is major shareholder of Insideout. Today a start-up based in Cairo, Egypt and focused on content management solutions for broadband and mobile networks.

Insideout10 is active in R&D via national and European projects and has established relationships in the academia and research sector via Sapienza Innovazione from Università La Sapienza, DIMA (Department of Information Engineering, Computer Science and Mathematics) at Università degli Studi dell'Aquila, The National Research Council in Italy (CNR) and the Salzburg Research Institute in Austria.

About MICO

MICO is a European Union part-funded research project to provide cross-media analysis solutions for online multimedia producers.

MICO develops models, standards and software tools to jointly analyse, query and retrieve information out of connected and related media objects (text, image, audio, video, office documents) to provide better information extraction results for more relevant search and information discovery.

About Greenpeace

Greenpeace is a non-governmental environmental organization with offices in over forty countries and with an international coordinating body in Amsterdam, the Netherlands. Greenpeace states its goal is to "ensure the ability of the Earth to nurture life in all its diversity" and focuses its campaigning on world wide issues such as climate change, deforestation, overfishing, commercial whaling, genetic engineering, and anti-nuclear issues. The global organization does not accept funding from governments, corporations, or political parties, relying on 2.9 million individual supporters and foundation grants.