

## Simposio 03: Digitalización, Datificación, e Investigación

# Datificación en los Archivos Digitales de los Medios de Comunicación: Nuevos Retos

Por Ángel Luis Garrido Marín<sup>1</sup> y Carlos Bobed Lisbona<sup>2</sup>

*(1) SID Research Group  
Departamento de Informática e Ingeniería de Sistemas  
Universidad de Zaragoza  
Zaragoza (España)  
Email: garrido@unizar.es*

*(2) SEMLIS Research Group  
IRISA / University of Rennes 1  
Rennes (Francia)  
E-mail: carlos.bobed-lisbona@irisa.fr*

### Palabras Clave (Tesauro SKOS):

- **Informatización de archivos** (<http://skos.um.es/unescothes/C00218>)
- **Industria de la información** (<http://skos.um.es/unescothes/COL520>)
- **Internet** (<http://skos.um.es/unescothes/C02091>)
- **Automatización de bibliotecas** (<http://skos.um.es/unescothes/C02279>)

## Introducción

En los últimos años los archivos digitales de los medios de comunicación han sufrido una gran transformación. Se ha pasado de usar métodos artesanales a su casi completa digitalización, lo que ha llevado a cambiar no sólo la forma de almacenar y clasificar la información, sino que también ha modificado su forma de acceso. Por otra parte, hemos asistido en la Web a un crecimiento exponencial de contenidos que pueden resultar también de gran ayuda en la elaboración de noticias.

El problema viene cuando un periodista está elaborando un nuevo contenido y necesita apoyo documental. A las habituales y numerosas informaciones que proporcionan las agencias se suman los contenidos propios, y toda la vasta información (y desinformación) que puede proporcionar Internet. Cómo encontrar en medio de este maremágnum digital la información más adecuada y cómo filtrar lo que es innecesario ó erróneo, es un reto diario para las cada día más menguadas redacciones de los medios de comunicación.

## Método

Desde nuestro grupo de investigación venimos trabajando desde hace varios años en sistemas inteligentes, que, sirviéndose de técnicas de procesamiento del lenguaje natural, de la semántica, y de los recursos automáticos que proporciona la Web (como el Linked Data), sean capaces de asistir en su trabajo diario a los usuarios de los archivos digitales. La datificación de estos archivos es ya un hecho y estas metodologías permiten su explotación con novedosos enfoques.

## Resultados y Discusión

Nuestros métodos y sistemas se han probado con resultados prometedores en entornos de pruebas y sobre entornos reales, siendo habitual nuestra colaboración con medios de HENNEO (antiguo Grupo Herald). Diversas publicaciones científicas a nivel internacional avalan nuestros resultados. Aunque queda todavía camino por recorrer a nivel de investigación, los primeros resultados empiezan a vislumbrarse y pronto estarán al alcance de cualquier medio de comunicación.

*Este trabajo de investigación ha sido financiado a través de los proyectos TIN2013-46238-C4-4-R y TIN2016-78011-C4-3-R (AEI/FEDER, UE)*

## Referencias

*Angel Luis Garrido and Susana Sangiao and Oscar Cardiel, "Improving the Generation of Infoboxes from Data Silos through Machine Learning and the use of Semantic Repositories", International Journal on Artificial Intelligence Tools 26(5), World Scientific, ISSN: 0218-2130, pp. 1-28, October 2017.*

*Angel Luis Garrido and Oscar Cardiel and Andrea Alexendri and Ruben Quilez, "Combining Machine Learning and Knowledge-Based Systems for Summarizing Interviews", 23rd International Symposium on Methodologies for Intelligent Systems (ISMIS 2017), Warsaw (Poland), Springer, ISBN 978-3-319-60437-4, ISSN 0302-9743, pp. 240-250, June 2017.*

*Angel Luis Garrido and Susana Sangiao and Oscar Cardiel and Adrian Gañan and Alejandro Bean and Sergio Ilarri, "NEREA: Named Entity Recognition and Disambiguation Exploiting Local Document Repositories", 28th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2016), San Jose (USA), IEEE Computer Society, ISBN 978-1-5090-4459-7, ISSN 2375-0197, pp. 1035-1042, November 2016.*

*Angel L. Garrido and Pilar Blazquez and Maria G. Buey and Sergio Ilarri, "Knowledge Obtenion Combining Information Extraction Techniques with Linked Data", 24th International World Wide Web Conference (WWW 2015), Florence (Italy), ACM, ISBN: 978-1-4503-3473-0, pp. 643-648, May 2015.*

*Angel Luis Garrido and Maria G. Buey and Sergio Ilarri and Eduardo Mena, "GEO-NASS: A Semantic Tagging Experience from Geographical Data on the Media", 17th East-European Conference on Advances in Databases and Information Systems (ADBIS 2013), Genoa (Italy), Springer Verlag LNCS, ISSN 0302-9743, ISBN 978-3-642-40682-9, volume 8133, pp. 56-69, September 2013.*

*Angel Luis Garrido and Oscar Gómez and Sergio Ilarri and Eduardo Mena, "An Experience Developing a Semantic Annotation System in a Media Group", 17th International conference on Applications of Natural Language Processing to Information Systems (NLDB 2012), Groningen (The Netherlands), Springer Verlag LNCS, ISSN 0302-9743, ISBN 978-3-642-31177-2, volume 7337, pp. 333-338, June 2012.*

---

### Ángel Luis Garrido Marín

*Ángel Luis Garrido es Máster en Ingeniería de Sistemas e Informática y Sistemas. Trabaja desde el 2001 en el mundo de los medios de comunicación, y está desde hace años de responsable corporativo de desarrollo en el grupo HENNEO. Colabora activamente en tareas de investigación en colaboración con el Grupo de Investigación de Sistemas de Información Distribuida (SID) de la Universidad de Zaragoza, dirigido por el Dr. Eduardo Mena Nieto.*

### Carlos Bobed Lisboa

*Carlos Bobed Lisboa es Doctor en Ingeniería Informática por la Universidad de Zaragoza, dónde realizó la tesis bajo la supervisión del Dr. Eduardo Mena Nieto. Actualmente es investigador postdoctoral en la Universidad de Rennes 1, en el grupo de investigación SEMMLIS dirigido por el Dr. Sébastien Ferré. Autor de numerosos artículos en revistas y conferencias internacionales, sus áreas de interés se centran principalmente en la Web Semántica y la Computación Móvil.*