

TextMess 2.0: Las Tecnologías del Lenguaje Humano ante los nuevos retos de la comunicación digital

TextMess 2.0: The New Digital Media Challenges Facing Human Language Technologies

Patricio Martínez-Barco
Univ. de Alicante (GPLSI – DLSI)
patricio@dlsi.ua.es

L. Alfonso Ureña López
Univ. de Jaén (SINAI)
laurena@ujaen.es

M. Antònia Martí
Universitat de Barcelona (CLiC)
amarti@ub.edu.es

Paolo Rosso
Univ. Pol. de València (ELiRF)
proso@dsic.upv.es

Resumen: El objetivo general de este proyecto se centra en el estudio, desarrollo y experimentación de diferentes técnicas y sistemas basados en Tecnologías del Lenguaje Humano (TLH) para el desarrollo de la próxima generación de sistemas de procesamiento inteligente de la información digital (modelado, recuperación, tratamiento, comprensión y descubrimiento) afrontando los actuales retos de la comunicación digital. En este nuevo escenario, los sistemas deben incorporar capacidades de razonamiento que descubrirán la subjetividad de la información en todos sus contextos (espacial, temporal y emocional) analizando las diferentes dimensiones de uso (multilingüedad, multimodalidad y registro).

Palabras clave: TLH, registros del lenguaje, minería de opiniones, análisis de sentimientos

Abstract: The overall aim of this project focuses on the study, development and experimentation with different techniques and systems based on Human Language Technologies (HLT) for developing the next generation of intelligent digital information processing systems (modelling, retrieval, processing, comprehension and detection), in order to meet the present challenges posed by digital media. In this new scenario, systems have to incorporate the reasoning capability to ascertain the subjectivity of information in all contexts (spatial, temporal and emotional), while analysing the various dimensional uses (multilingualism, multimodality and register).

Keywords: HLT, language registers, opinion mining, sentiment analysis.

1 Descripción

La actual Sociedad de la Información ha variado la forma en la que se produce la interacción comunicativa entre los humanos. A diferencia de la comunicación convencional (prensa escrita, radio, o televisión), los nuevos instrumentos para la propagación de la información permiten que sus usuarios no sean sólo consumidores, sino también productores de sus propios mensajes.

Estos nuevos mensajes se integran día a día en Internet conformando la llamada web 2.0 como la mayor plataforma comunicativa y colaborativa de la historia. Como consecuencia, se dispone de un gran volumen de información digital en el que se combinan múltiples lenguas (multilingüedad), formatos de representación (multimodalidad), y registros de uso con

diferentes grados de ritualidad (formal/informal) y especialización, emitidos desde diferentes actitudes subjetivas y desde diferentes situaciones de espacio y tiempo que deben ser analizadas y comprendidas por sistemas inteligentes para su adecuada explotación por organizaciones y particulares.

La finalidad del proyecto TEXTMESS 2.0¹ es, por tanto, el estudio, desarrollo y experimentación de diferentes técnicas y sistemas basados en Tecnologías del Lenguaje Humano (TLH) para el desarrollo de la próxima generación de sistemas de procesamiento inteligente de la información digital (modelado, recuperación, tratamiento, comprensión y descubrimiento) afrontando los actuales retos de la comunicación digital. En este nuevo

¹ <http://intime.dlsi.ua.es/textmess20>

escenario, los sistemas deberán incorporar capacidades de razonamiento que descubrirán la subjetividad de la información en todos sus contextos (espacial, temporal y emocional) analizando las diferentes dimensiones de uso (multilingüedad, multimodalidad y registro).

TextMess 2.0 está financiado por el MICINN (TIN2009-13391-C04) con el apoyo de la acción complementaria Araknion (FFI2010-114774E), así como por cofinanciación de los proyectos PROMETEO/ 2009/119, ACOMP/2010/286 y ACOMP/2011/ 001 de la Generalitat Valenciana, la red Tratamiento de Información Multimodal y Multilingüe (TIN2009-06135-E), el proyecto de excelencia de la Junta de Andalucía GeOasis (P08-TIC-4199), Marie Curie actions WiQ-Ei (269180) y proyecto VLC/CAMPUS microcluster MIIS.

2 *Objetivos*

Con el planteamiento anterior, los objetivos generales que presenta el proyecto coordinado son los siguientes:

- Afrontar los nuevos retos de la comunicación digital desde la creación, adaptación y mejora de recursos, técnicas y herramientas que modelan el lenguaje digital actual mediante aproximaciones de aprendizaje automático, donde se estudiarán las implicaciones de los diversos registros de uso, la multilingüedad, la multimodalidad, y los contextos espaciales, temporales y emocionales.
- Desarrollar sistemas inteligentes de procesamiento (recuperación, tratamiento, comprensión y descubrimiento) de la información adaptados a las nuevas formas de comunicación con capacidad de interpretar y valorar el contexto del mensaje.
- Integración de los recursos, herramientas y sistemas de TLH desarrollados y recopilados conformando un metasistema de procesamiento inteligente de información digital, cuya validez se demostrará sobre un escenario de uso concreto de inteligencia competitiva en el entorno socio-económico.
- Participar y promover actividades relacionadas con las líneas de investigación del proyecto en campañas, congresos, talleres, seminarios y redes temáticas.

3 *Proyecto coordinado*

Para la consecución del objetivo global y el desarrollo óptimo de las diferentes líneas de actuación del proyecto descritas anteriormente,

se ha propuesto la coordinación de cuatro subproyectos complementarios cuyos objetivos específicos abarcan los objetivos globales planteados, y cuya reunificación proporciona el valor añadido que se buscaba en la coordinación.

En primer lugar, es necesario el modelado de la información en todos sus registros. Éste es el objetivo principal del subproyecto TEXT-Knowledge 2.0 que se realiza desde el equipo de investigación de la Universitat de Barcelona. Por otra parte, es necesario obtener metodologías y técnicas para la recuperación y el tratamiento de la información, conformando el objetivo del subproyecto TEXT-COOL 2.0 que lleva a cabo el equipo de la Universidad de Jaén. En tercer lugar, se debe conseguir la comprensión de la información y su aplicación a las necesidades de negocio actual, objetivo del subproyecto TEXT-Enterprise 2.0 que se desarrolla desde el equipo de la Universitat Politècnica de València. Por último, el descubrimiento y síntesis de la información subyacente, así como la integración de todas estas tecnologías en un marco global de procesamiento inteligente de la información digital de las fuentes 2.0, es el objetivo del subproyecto TEXT-MESS 2.0 del equipo de la Universidad de Alicante. Así cada subproyecto se centra en el estudio y desarrollo de técnicas en cada uno de los grandes bloques anteriores colaborando estrechamente para lograr el objetivo común propuesto.

4 *Beneficios del proyecto*

Como contribuciones científicas esperables del proyecto se espera crear conocimiento en forma de herramientas y corpus etiquetados, disponibles para los demás miembros de la comunidad científica, no sólo en publicaciones en congresos y revistas relevantes, sino también como resultados materiales del proyecto, con lo que se busca aumentar el grado de interactividad con sus posibles usuarios.

Por otra parte, las contribuciones tecnológicas esperables se concretan en el desarrollo de sistemas de RI, EI, BR, producción de resúmenes y detección de plagios con capacidad de adaptación para diferentes escenarios multilingües, multimodales y de múltiples registros. Además, el proyecto está construyendo un metasistema que integra los diferentes sistemas y tecnologías de TLH desarrolladas.