



CAEPIA-ESTYLF 2018

SESIÓN ESPECIAL

“SOFT COMPUTING Y GENERACIÓN DE LENGUAJE NATURAL”

Organizadores: Alberto Bugarín, Nicolás Marín, Alejandro Ramos, Daniel Sánchez

El objetivo principal de la sesión es reunir a grupos de investigación españoles de los ámbitos “Soft Computing” y “Generación de Lenguaje Natural (NLG)” para discutir acerca de los posibles puntos de convergencia de los dos ámbitos, en un sentido amplio.

Por un lado, desde el Soft Computing, las principales aproximaciones al ámbito de la NLG se concretan en las “Descripciones Lingüísticas de Datos (LDD)”, en las que se aplican métodos y técnicas de la “computación con palabras” para generar representaciones lingüísticas que contienen la información más relevante de un conjunto de datos y en las que se incluyen términos imprecisos. Las LDD suelen basarse en protoformas y, en general, no han atendido a todas las etapas que habitualmente se consideran generalmente en los sistemas NLG.

Por su parte, desde la NLG, los sistemas que producen descripciones o narrativas a partir de datos (“sistemas Data-To-Text (D2T)”) cuentan con una metodología de desarrollo ampliamente aceptada, que se ha venido aplicando por mucho tiempo en ámbitos diversos. Las narrativas generadas incluyen con frecuencia términos y expresiones de naturaleza vaga, pero no suelen utilizar aproximaciones basadas en Soft Computing para su modelado.

Hasta hace poco, las LDD y la NLG/D2T habían avanzado en sus propuestas, metodologías y desarrollos de forma paralela, a pesar de que el objetivo final de los dos ámbitos es producir narrativas expresadas en lenguaje natural dirigidas a usuarios finales que las van a consumir. Sin embargo, en los últimos años se ha creado un entorno de colaboración entre ambos campos, tanto a nivel nacional como internacional. Prueba de ello es la reciente concesión de la Red Nacional de Excelencia En Tratamiento Inteligente De Datos y Generación De Lenguaje Natural, que aúna a los principales grupos españoles de ambas disciplinas.

En este contexto, la sesión propuesta se erige como la primera de las actividades principales de interés para los grupos que conforman la mencionada Red. Al mismo tiempo, expande el alcance de la sesión homónima acontecida en el ESTYLF 2016, germen a su vez de la actual Red de Excelencia. Así pues, esta sesión se dirige hacia grupos de investigación de los dos ámbitos, con el objeto de presentar las diferentes aproximaciones y visiones complementarias procedentes de ambos paradigmas, para tratar de continuar definiendo posibles sinergias o puntos de encuentro de interés común.

La sesión pretende suscitar debate y discusión acerca de, entre otros, los siguientes puntos:

- las necesidades e interés que desde la NLG (y, en particular, los sistemas D2T) se tiene en cuanto al modelado de términos y expresiones con incertidumbre
- en qué medida el “Soft Computing” y las LDD pueden atender dichas necesidades y su rol en el proceso NLG
- el papel de otras técnicas de IA en general (no específicamente las basadas en “Soft Computing”) en las fases de dicho proceso NLG
- aplicaciones reales de LDD y NLG
- ...además de cualquier otro tema o punto de encuentro que pudiera ser de interés común a ambas aproximaciones

Esta sesión continua la edición homónima organizada en el ESTYLF 2016, así como la serie de sesiones especiales en la temática “Descripciones lingüísticas de datos”, organizadas por algunos de los organizadores, en eventos internacionales como WCCI 2018, FUZZ-IEEE 2017, IPMU 2016, WCCI 2016, o IFSA-EUSFLAT 2015, entre otros.

La lista (no exhaustiva) de tópicos de interés para la sesión es, necesariamente, interdisciplinar y procedente de los dos ámbitos cuya convergencia se busca promover: Soft Computing y Generación de Lenguaje Natural:

- Gestión de la vaguedad e incertidumbre en Generación de Lenguaje Natural.
- Descripciones Lingüísticas de Datos: modelos.
- Modelado de términos, vocabularios o expresiones.
- Generación de expresiones de referencia sobre propiedades borrosas.
- "Data mining" para determinación de contenidos u otras fases del proceso NLG.
- Sistemas de Generación de Lenguaje Natural y Sistemas "Data-To-Text".
- Retos de la IA y el Soft Computing para la Generación de Lenguaje Natural.
- Evaluación de la calidad de las narrativas en LDD y NLG.
- Aplicaciones de sistemas basados en LDD y/o NLG.
- Hibridación entre el Soft Computing y la NLG.
- Generación de descripciones textuales a partir de datos semánticos (RDF).
- Generación de resúmenes sobre grafos de conocimiento (Knowledge Graphs).
- Curación supervisada de datos a partir de sus descripciones textuales.
- Uso de Linked Open Data (DBpedia, GraphDB, etc.) en NLG y/o Soft Computing.
- Accesibilidad y NLG.

El enfoque o contenido de los trabajos puede incluir desarrollos, proyectos, aplicaciones o actividades de investigación recientes en cualquiera de las dos áreas, "position papers" acerca de las respectivas visiones o retos existentes desde cada uno de los dos paradigmas, trabajos de I+D+i en progreso de los dos ámbitos o, en general, cualquier otro enfoque que promueva la presentación de retos o temas actuales de I+D+i en cualquiera de los dos ámbitos que favorezcan la discusión conjunta y la interacción con el resto de grupos participantes en la sesión.

Fechas importantes:

- Envío de trabajos: 20 de junio de 2018 (Fecha firme)
- Notificación de aceptación: 20 de julio de 2018
- Versión Camera-ready: 10 de septiembre de 2018
- Inscripción reducida: 10 de septiembre de 2018
- Celebración del Congreso: 23-26 de octubre de 2018

Instrucciones para autores

Las instrucciones detalladas para los autores, formato de los trabajos y modo de envío están disponibles en la Web del Congreso:

<http://sci2s.ugr.es/caepia18/estylf/index.html>

Se aceptan dos tipos de trabajos:

1. Trabajos inéditos que podrán publicarse en español o inglés. La extensión de estos trabajos será de hasta 6 páginas, siguiendo el estilo IEEE.
2. Trabajos aceptados recientemente (key works) en revistas de prestigio. Su extensión será de hasta 2 páginas, siguiendo el estilo IEEE y clara referencia al trabajo ya publicado.

Información de Contacto:

Alberto Bugarín Centro de Investigación en Tecnoloxías da Información (CITIUS) Universidade de Santiago de Compostela 15782, Santiago de Compostela Email: alberto.bugarin.diz@usc.es	Nicolás Marín Departamento de Ciencia de la Computación e Inteligencia Artificial Universidad de Granada 18071, Granada Email: nicm@decsai.ugr.es
Alejandro Ramos Centro de Investigación en Tecnoloxías da Información (CITIUS) Universidade de Santiago de Compostela 15782, Santiago de Compostela Email: alejandro.ramos@usc.es	Daniel Sánchez Departamento de Ciencia de la Computación e Inteligencia Artificial Universidad de Granada 18071, Granada Email: daniel@decsai.ugr.es