

[Inicio](#) > El investigador del BSC Aitor Gonzalez-Agirre recibe el premio a la mejor tesis de 2017 en el congreso SEPLN

---

## [El investigador del BSC Aitor Gonzalez-Agirre recibe el premio a la mejor tesis de 2017 en el congreso SEPLN](#)

"*Computational Models for Semantic Textual Similarity*" pretende avanzar en los modelos computacionales para la evaluación del significado de las oraciones.



Aitor Gonzalez-Agirre, investigador del grupo de Text Mining del BSC, ha recibido el premio a la mejor tesis del 2017 en el [Congreso SEPLN](#) (Sociedad Española para el Procesamiento del Lenguaje Natural) celebrada en Sevilla del 19 al 21 de septiembre. Este congreso tiene como objetivo ofrecer un foro de debate y comunicación donde la comunidad científica y la empresa puedan presentar los trabajos de investigación y los hallazgos más recientes en el área del Procesamiento del Lenguaje Natural (PLN).

La tesis de Gonzalez-Agirre, "*Computational Models for Semantic Textual Similarity*", tiene como objetivo avanzar en los modelos computacionales para la evaluación del significado de las oraciones. Para lograr este objetivo, define dos tareas y desarrolla sistemas de última generación que abordan ambas tareas: Similitud Semántica Textual (STS, de sus siglas en inglés) y Similitud Tipada (*Typed Similarity* en inglés). STS tiene como objetivo medir el grado de equivalencia semántica entre dos oraciones asignando valores de similitud gradual que capturan los tonos intermedios de similitud. Gonzalez-Agirre ha recogido pares de oraciones para construir conjuntos de datos para STS, un total de 15.436 pares de oraciones, siendo de lejos la mayor colección de datos para STS. El investigador también ha diseñado, construido y evaluado un nuevo enfoque para combinar los métodos basados ??en conocimiento y en corpus utilizando un cubo. Este nuevo sistema para STS está a la par con los enfoques de última generación que hacen uso de Aprendizaje Automático o *Machine Learning* (ML) sin usar nada de eso, aunque permite emplear ML, mejorando los resultados. La tarea Similitud Tipada intenta identificar el tipo de relación que existe entre pares de artículos de patrimonio cultural de una biblioteca digital. Proporcionar una razón por la cual los artículos son similares tiene aplicaciones en recomendación, personalización y búsqueda. Se identificaron una variedad de tipos de similitud en esta colección y se anotó un conjunto de 1.500 pares de artículos de la colección mediante crowdsourcing. Finalmente, presenta sistemas capaces de resolver la tarea de Similitud Tipada. El mejor de estos sistemas se empleó en un entorno real para recomendar artículos similares a los usuarios de una biblioteca digital en línea.

### **Sobre el grupo de Text Mining del BSC**

El [grupo de Biological Text Mining](#) se centra en la aplicación y desarrollo de tecnologías de minería de texto biomédica, que se están convirtiendo en una herramienta clave para la explotación eficiente de la información contenida en repositorios de datos no estructurados incluyendo literatura científica, registros electrónicos de salud (EHR), patentes, biobanco metadatos, ensayos clínicos y redes sociales. La unidad tiene un interés particular en el procesamiento de documentos clínicos escritos en español y otros idiomas cooficiales en el área de temas relacionados con la salud y la integración de información molecular y biológica derivada de la literatura. La unidad está totalmente financiada a través del "Plan de Impulso de las Tecnologías del Lenguaje de la Agenda Digital (PITL)", en el marco de un acuerdo ("encomienda") entre el Secretario de Estado de Telecomunicaciones del Ministerio de Energía español, Turismo y Agenda Digital (MINETAD) y CNIO.







Barcelona Supercomputing Center - Centro Nacional de Supercomputación

---

**Source URL (retrieved on 26 Dic 2024 - 02:11):** <https://www.bsc.es/es/noticias/noticias-del-bsc/el-investigador-del-bsc-aitor-gonzalez-agirre-recibe-el-premio-la-mejor-tesis-de-2017-en-el-congreso>