

[Inici](#) > L'investigador del BSC Aitor Gonzalez-Agirre rep el premi a la millor tesi de 2017 en al congrés SEPLN

L'investigador del BSC Aitor Gonzalez-Agirre rep el premi a la millor tesi de 2017 en al congrés SEPLN

Computational Models for Semantic Textual Similarity" pretén avançar en els models computacionals per a l'avaluació del significat de les oracions.



Aitor Gonzalez-Agirre, investigador del grup de Text Mining del BSC, ha rebut el premi a la millor tesi del 2017 al [Congrés SEPLN](#) (Societat Espanyola per al Processament del Llenguatge Natural) celebrada a Sevilla del 19 al 21 de setembre. Aquest congrés té com a objectiu oferir un fòrum de debat i comunicació on la comunitat científica i l'empresa puguin presentar els treballs de recerca i les troballes més recents a l'àrea del Processament del Llenguatge Natural (PLN).

La tesi de Gonzalez-Agirre, "*Computational Models for Semantic Textual Similarity*", té com a objectiu avançar en els models computacionals per a l'avaluació del significat de les oracions. Per aconseguir aquest objectiu, defineix dues tasques i desenvolupa sistemes d'última generació que aborden les dues tasques: Similitud Semàntica Textual (STS, de les seves sigles en anglès) i Similitud Tipada (*Typed Similarity* en anglès). STS té com a objectiu mesurar el grau d'equivalència semàntica entre dues oracions assignant valors de similitud gradual que capturen els tons intermedis de similitud. Gonzalez-Agirre ha recollit parells d'oracions per construir conjunts de dades per STS, un total de 15.436 parells d'oracions, sent de lluny la major col·lecció de dades per STS. L'investigador també ha dissenyat, construït i avaluat un nou enfocament per combinar els mètodes basats en coneixement i en corpus utilitzant un cub. Aquest nou sistema per STS està a la par amb els enfocaments d'última generació que fan ús d'Aprenentatge Automàtic o Machine Learning (ML) sense usar res d'això, encara que permet emprar ML, millorant els resultats. La tasca Similitud Tipada intenta identificar el tipus de relació que hi ha entre parells d'articles de patrimoni cultural d'una biblioteca digital. Proporcionar una raó per la qual els articles són similars té aplicacions en recomanació, personalització i cerca. Es van identificar una varietat de tipus de similitud en aquesta col·lecció i es va anotar un conjunt de 1.500 parells d'articles de la col·lecció mitjançant crowdsourcing. Finalment, presenta sistemes capaços de resoldre la tasca de Similitud Tipada. El millor d'aquests sistemes es va emprar en un entorn real per recomanar articles similars als usuaris d'una biblioteca digital en línia.

Sobre el grup de Text Mining del BSC

El [grup de Biological Text Mining](#) se centra en l'aplicació i desenvolupament de tecnologies de mineria de text biomèdica, que s'estan convertint en una eina clau per a l'explotació eficient de la informació continguda en repositoris de dades no estructurats incloent literatura científica, registres electrònics de salut (EHR), patents, biobanc metadades, assaigs clínics i xarxes socials. La unitat té un interès particular en el processament de documents clínics escrits en espanyol i altres idiomes cooficials en l'àrea de temes relacionats amb la salut i la integració d'informació molecular i biològica derivada de la literatura. La unitat està totalment finançada a través del "Pla d'Impuls de les Tecnologies del Llenguatge de l'Agenda Digital (PITL)", en el marc d'un acord ("encomana") entre el Secretari d'Estat de Telecomunicacions del Ministeri d'Energia espanyol, Turisme i Agenda Digital (MINETAD) i CNIO.





Barcelona Supercomputing Center - Centro Nacional de Supercomputación

Source URL (retrieved on 16 set 2024 - 11:59): <https://www.bsc.es/ca/noticies/noticies-del-bsc/linvestigador-del-bsc-aitor-gonzalez-agirre-rep-el-premi-la-millor-tesi-de-2017-en-al-congr%C3%A9s-seplin>