

TELEPARES: Tecnologías de la lengua para análisis de opiniones en redes sociales



Descrición

Cada vez é maior o número de usuarios que empregan os medios web de comunicación social baseados en microtextos para compartir as súas opinións e experiencias achega de produtos, servizos ou persoas. Esta situación espertou un gran interese polo desenvolvemento de plataformas informáticas que procesen estes textos e exploten o seu potencial para coñecer o punto de vista dunha ampla base de consumidores, expresado en diferentes linguas, sobre unha gran variedade de temas. A este respecto, a análise de sentimento ou minería de opinións, é unha recente área de investigación centrada en determinar automaticamente si nun texto opínase ou non, si a polaridad ou sentimento que se expresa nel é positiva, negativa ou mixta; e en extraer automaticamente a percepción dun autor sobre aspectos concretos dun tema. Con todo, as solucións actuais están moi limitadas polo seu escaso recurso ás tecnoloxías da lingua, ao basearse nun procesado superficial que non ten en conta as relacións sintácticas entre palabras nin as súas roles semánticos nas oracións, o cal resta capacidade de comprensión nuns textos xa de seu exiguos. Ademais, a maioría destas ferramentas adoptan ao inglés como lingua basee, coa consiguiente vantaxe para usuarios, organizacións e empresas de países angloparlantes.

Aínda que os microtextos presentan características léxicas e sintácticas propias respecto á linguaxe estándar, certos aspectos básicos da linguaxe han de ser respectados para resultar lexibles. Neste proxecto propoñemos explotar este feito para obter unha mellora do soporte lingüístico integrado no tratamento de microtextos no noso ámbito de interese natural, o español e o galego, linguas para as que recentes informes europeos poñen de manifesto un importante déficit, cunha especial incidencia na carencia de recursos sintácticos. Para conseguilo será necesario combinar a experiencia e coñecementos dos equipos participantes en Lingüística Computacional, Recuperación de Información e Adquisición do Coñecemento.

A meta final consiste en desenvolver un sistema efectivo de análise de opinións en español e galego para as redes sociais baseadas en microtextos. Para iso será preciso mellorar o rendemento das técnicas actuais de análises sobre texto estándar, deseñar mecanismos de adaptación a microtextos daqueles modelos e métodos de análises que son máis efectivos en linguaxe estándar; e realizar unha proxección de modelos, métodos e recursos efectivos noutras linguas.

INVESTIGADORES

Proxecto de

Compiladores y Lenguajes (COLE), Universidade de Vigo

Investigador principal externo

Manuel Vilares Ferro

Colaboradores do CiTIUS

Pablo Gamallo Otero

DETALLES

Data de execución:

06/10/2015 - 31/12/2017

Financiado por

Programa Estatal de I+D+i Orientada a los Retos de la Sociedad, Ministerio de Economía y Competitividad, FFI2014-51978-C2-1-R



PO FEDER Galicia 2014-2020 "Unha maneira de facer Europa"

PUBLICACIÓNS

Strategies for building high quality bilingual lexicons from comparable corpora

Parallel Corpora for Contrastive and Translation Studies: New resources and applications, 2019

Explorando métodos non-supervisados para calcular a similitude semántica textual

Linguamática, 2019

Using the Outlier Detection Task to Evaluate Distributional Semantic Models

Machine Learning and Knowledge Extraction, 2018

[Ver todas](#)

PROGRAMAS CIENTÍFICOS

Tecnoloxías da Linguaxe Natural