

uComp research project delivers first results under open source license

June 28 2013

Methods to extract knowledge from social media intelligently and automatically are currently being developed at MODUL University Vienna - and the latest advances have just been published in preparation of an international conference. These advances come in the form of an open source tool to collect and process publicly available social media information.

The tool supports text acquisition, [language recognition](#), detection of phonetic similarities, as well as the standardized integration and archiving of the captured information. The [open source tool](#) represents a major step forward in the uComp project of MODUL University Vienna (Austria) and its European partners. Using the domain of [climate change](#) as an example, the project combines cutting-edge methods to automatically capture information from complex sources and combine it with collective human intelligence in the tradition of the "wisdom of the crowds".

The internet is very different from a well-structured database. Unlike libraries or large corporate archives, online information is fragmented and disordered, which makes it difficult to extract knowledge automatically. The emergence of social media has further complicated the process. It is difficult to determine the specific context of a posting, and the use of slang, dialects or foreign words challenges existing tools for text analysis. Scientists and researchers are currently working on solving this problem in the uComp project jointly conducted by MODUL University Vienna and partner organizations from Austria,

England and France. After only six months, first results have now been published in preparation of the 7th International Conference for Knowledge Capture (K-Cap 2013) in Banff, Canada.

Man/machine symbiosis

The objective of uComp is explained by the head of the Department of New Media Technology at MODUL University Vienna, Prof. Arno Scharl, using the domain of climate change as a use case: "Millions of people express their opinions in social media, but with conventional methods we are unable to determine the collective mood expressed in social media in real time. We do not know which aspects move people, mobilize people or stimulate their thoughts. The technologies from the uComp project provide us with better ways to capture opinions - on a global basis, irrespective of language barriers, national borders and cultural differences."

The key aspect of uComp for Prof. Scharl, who also serves as the project's Technical Director, is the combination of collective human intelligence and automated knowledge extraction by software tools. The first step to achieving this vision has successfully been taken with the "extensible Web Retrieval Toolkit" (eWRT), which has now been published in a scientific paper. As an open source tool, eWRT promotes a transparent approach to analyzing data from [social media](#) platforms. The system captures data from many different public sources and accurately identifies the language of the gathered information items. Additional functions include the ability to archive large volumes of data, including the management and normalization of relevant metadata (= data that describes the structure and content of documents).

The next two-and-a-half years will focus on using collective [human intelligence](#) for the analysis and validation of data gathered with eWRT. Games with a purpose represent a promising approach in the field of

human computation (HC). Examples include online games for classifying documents or for evaluating automatic translations. By aiming to integrate such games into a comprehensive framework to identify complex knowledge patterns, the uComp project is entering unknown digital territory. As Prof. Scharl explains, "We are currently investigating ways of engaging people and providing incentives for participants to share their knowledge. At the same time we need to evaluate the reliability of their contributions, prevent manipulation and assess the quality of results. The uComp project will advance the state of the art by offering all these capabilities in an integrated, reusable framework."

The uComp project and the collaboration of Prof. Scharl's team with fellow researchers from England, France and Austria continue a successful tradition. The DIVINE project, funded by the Austrian Research Promotion Agency (FFG) and the Austrian Ministry for Transport Information and Technology (BMVIT), has already addressed important aspects on the dynamic integration and visualization of information spaces and made major contributions to the development of the eWRT software package.

More information: www.ucomp.eu/

Provided by University of Vienna

Citation: uComp research project delivers first results under open source license (2013, June 28) retrieved 27 April 2024 from <https://phys.org/news/2013-06-ucomp-results-source.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--