

System filters the arguments in texts and checks their quality

December 22 2016, by Jutta Witte



Quality check on sources: Prof. Iryna Gurevych (left) and her team. Credit: Sandra Junker

There is a sea of information and argumentation on the Internet covering every possible world-shattering subject. The Ubiquitous Knowledge

Processing Lab at TU Darmstadt is developing tools for a quality check.

Should [plastic water bottles](#) be banned? Has India the potential to become a world power? Should children be allowed to use mobile phones at school? You can find supporting arguments and help in decision-making from experts and non-experts on the net for virtually every controversy nowadays. But what about the quality of these technical texts, the information provided by interested amateurs and the contributions to debates? "So far, the vast majority of them are not validated", explains Professor Iryna Gurevych, [Head of the Ubiquitous Knowledge Processing \(UKP\) Lab at TU Darmstadt](#). So she and her research team are developing software instruments designed not only to filter the arguments out of the texts, but also to check their quality.

With the aid of the debate about the use of mobile phones, the "Digital Humanities" expert explains a possible scenario, in which a learning machine identifies arguments in a specific collection of documents and analyses them along with the associated corroborations should parents encourage their children to limit their [mobile phone](#) use?

A mother takes to the net looking for answers. The analysis system recognises the topic and uses keywords – such as children, parents, mobile phone or radio frequency emissions – to find the text fragments relevant to the query. It then starts a so-called predicate argument analysis in the individual fragments, which means that it is looking for the deed and its frame of reference in the sentences.

After the analysis of the individual fragments, references with regard to content are created between all the text passages found, and it uses its own knowledge database, as well as feedback from users who have commented on the relevant texts on the net, to identify premises, assertions and supporting or contradictory corroborations for the particular argument. The evaluation follows this categorisation.

Recognises shortcomings in the argumentation

If, for example, a statement only relates to "some study or other", i.e. its sources are extremely vague, or if it is one-sided and only contains supporting corroboration, the system will recognise these shortcomings in the argumentation. At the end, the anxious mother is presented with a graphic, showing whether the respective arguments are plausible and credible.

This is one of many application ideas for the automated analysis of argumentation on which the research scientists of the UKP Lab are currently working. Gurevych is convinced that this kind of quality checking would be a great step forward in this complex research field.

What the experts really need is all sorts of quality-checked training data to feed into their systems, so that they can develop the necessary algorithms, methods and ultimately also prototypes for the new generation of search engines. They have just created a database called "UKPConvArg2". The new collection encompasses around 9,000 pairs of arguments from social media discussions that people have evaluated with regard to quality, and coded. In each case, they are the arguments for and against a socially relevant topic.

"This database, which we are making available to the scientific community, does not only show which arguments are convincing and why. It also forms the basis for developing new methods for the empirical analysis of text data from the Internet", explains Ivan Habernal, a research scientist at the UKP Lab.

"It allows us to start a new discussion about the possibilities of machine learning." The UKP experts estimate that simple applications, such as segmenting texts into argumentative and non-argumentative parts within clearly defined text types, will be achievable in the near future – perhaps

as an additional tool for Google Search.

Writing assistants, to discover the weaknesses in the arguments in a student essay, for example and include them in a scoring system for exams, are also already technically feasible. But the major challenge is still the editing of texts from heterogeneous sources ranging from scientific papers to articles on social media.

Firstly, because creating the training data is a highly complex task, and secondly because the methods of analysis are developed for a specific type of text on the basis of this data, and can hardly be transferred, as things stand at the moment. "We have yet to resolve the question of scaling", says Gurevych. "This is a research task for the next five years." The "ArgumenText" project has just been given a positive funding recommendation, and the UKP experts want to use it to look into this question and transpose to new applications the tools for the automated analysis of assertions and corroborations that have already been used successfully in a specific application context.

Potential users of the new analysis instruments being produced in the UKP Lab include not only educationalists looking for help in correcting exam papers, but also companies who want to evaluate customer reports on their products, or journalists wanting to quickly and comprehensively research the viewpoints of the different factions involved in the latest "hot topics". The humanities and social sciences can also benefit from this, for example, when it is a matter of collecting and evaluating all the relevant text data that could substantiate or refute a theory.

"Nowadays it is not possible to do this by hand, because we simply do not have the capacity", says the IT specialist. The aim would be to create a tool to tap into and pre-structure the vast amounts of information from the different channels. But the interpretation of this machine-generated knowledge still remains the province of man. Gurevych and Habernal

stress that a machine could process vast amounts of data according to specific patterns, without getting tired. "But it will always lack the world knowledge of mankind, which is indispensable for categorising and transferring new information in larger contexts."

Provided by Technische Universität Darmstadt

Citation: System filters the arguments in texts and checks their quality (2016, December 22)
retrieved 3 May 2024 from <https://phys.org/news/2016-12-filters-arguments-texts-quality.html>

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.