

# Mining for meaning: Getting computers to understand natural language texts

July 18 2013

---

Programs that can understand language and can identify meaningful links between the various parts of a text is the focus of work being carried out in Saarbrücken by researchers like Ivan Titov. The computer scientist is currently developing a procedure that will enable computers to learn to identify semantically relevant relationships within texts. This research could mean that in future we will be able to ask our computer specific questions about the content of a text. The computer would then analyse the text and supply the user with the right answers.

Every student who has ever written a homework assignment or an academic essay is familiar with the problem: before you can start to write anything yourself, you usually have to battle through numerous texts and through pages and pages of references and academic literature. A [computer program](#) that can quickly process the text, provide a meaningful summary of its content or even answer questions about it would obviously be of great practical value in such a situation.

Ivan Titov and his team of researchers, splitting their time between Saarland University and the University of Amsterdam, are currently working on this problem. Titov is interested in how computers can learn to understand the meaning and the relationships between words in sentences and within texts. 'The model that we have developed simulates how humans create texts. In order to understand texts, we get our computers to work through this process but in the reverse direction: given the text the computer will uncover its meaning or even intent of the writer' explained Dr Titov. However, Titov and his group do not

themselves stipulate a fully detailed model and the rules contained within it, instead, they use millions of sentences to generate both the model and the rules. The sentences that are analysed are drawn from large collections, such as Wikipedia. Analysis of this massive dataset requires a lot of computing power, with the specially developed algorithms running on around one hundred computers.

The idea is to develop software that enables computers to identify hidden, context-dependent relationships between words and clauses in texts, as the following example shows. Looking at the two sentences: 'John has just graduated from Saarland University. He is now working for Google.' it is clear even to a computer that John and Saarland University are linked by the relationship 'has graduated' and that John and Google are connected by the relationship "is working for". But the model developed by the Saarbrücken computer scientists can also recognize that John studied at Saarland University; very probably in the Department of Computer Science and Informatics. Once computers can understand these patterns in human language, the next step for the researchers is to apply the method to get machines to automatically produce meaningful summaries of short texts and to answer questions about the text content.

Beside Ivan Titov Hans Uszkoreit is awarded with a Google Focused Award worth US\$ 220.000. Uszkoreit is professor of Computational Linguistics at Saarland University and Scientific Director at the German Research Center for Artificial Intelligence. He is interested in how computers can identify linguistic relationships in large text collections.

Through its Focused Research Award program, the search engine provider Google supports research of major interest to the company itself and to the field of informatics. Prize winners receive free access to Google tools and technologies.

Provided by Saarland University

Citation: Mining for meaning: Getting computers to understand natural language texts (2013, July 18) retrieved 12 May 2024 from <https://phys.org/news/2013-07-natural-language-texts.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.