

NEWS, but not as we know it

April 18 2006

It will mean stories can be defined, on the fly, with a precision greater than a library's card catalogue. The News Engine Web Services (NEWS) platform is aimed at news agencies, governments and large enterprises and will enable them to develop highly advanced analysis to raw text, with a vast number of potential applications.

News agencies will be able to automatically create very highly personalised news profiles for readers. Governments will be able to analyse social and political trends through newspaper reports, at a much higher level of detail than was possible previously, and large businesses will be able to study market and product developments.

The project that developed the platform even managed to develop a proof-of-concept service for analysing audio, by combining their system with a commercial voice recognition programme.

At the heart of this functionality is the powerful classification and ontology-based annotation system that can work across languages. "News classifications up to now typically consisted of about 12 terms, like sport, world news, finance, that a journalist knew off by heart," says Dr Ansgar Bernardi, deputy head of the Knowledge Management Group at DFKI, the German Research Centre for Artificial Intelligence, and coordinator of the IST-funded NEWS project.

"That's not very precise. Our system can automatically analyse a story and access 1300 classification terms to define it," says Bernardi.

What's more it can access a large ontology of terms related to the specific story definitions within a class, terms like president, head-of-state and government in the politics class, for example. The end result is a very large data set of standardised terms that define the story's content.

That data set can then be used in a huge variety of ways to potentially answer almost any query a user can imagine. A simple example: "Show me news items about the US president in January 2006" will deliver news items about George W. Bush in this time frame.

"We expect that platform users will take the basic functionality and develop around it to respond to the information they want to analyse," says Bernardi. The system also needs to be 'trained' for analysis of specific topics.

To avoid 'false positives', where two people of the same name are confused, for example, or where two cities have the same name, the NEWS team developed IdentityRank, an adaptive algorithm for instance disambiguation.

"It really started out as a by-product of our main work, but it works well and I think it may generate quite a bit of scientific interest," says Bernardi.

It's only one of NEWS' many achievements, and work will not stop there. "We have developed a great network during the project and the consortium has agreed to offer mutual support for a further two years. In the meantime we are pursuing commercial opportunities, several news agencies are interested in the platform, and we had a lot of exposure at CEBIT '05 and '06," says Bernardi.

Source: [IST Results](#)

Citation: NEWS, but not as we know it (2006, April 18) retrieved 9 April 2024 from <https://phys.org/news/2006-04-news.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.