

Sorting facts and opinions for Homeland Security

September 24 2006

What are newspapers around the world saying about the latest speech by President George W. Bush? More importantly, how much of what they are saying is factual and how much opinion? And down the line, are some of the opinions being presented as if they were facts?

A new research program by a Cornell computer scientist, in collaboration with colleagues at the University of Pittsburgh and University of Utah, aims to teach computers to scan through text and sort opinion from fact. The research is funded by the U.S. Department of Homeland Security, which has designated the consortium of three universities as one of four University Affiliate Centers (UAC) to conduct research on advanced methods for information analysis and to develop computational technologies that contribute to national security. Cornell will receive \$850,000 of \$2.4 million in funding provided for the consortium over three years.

"Lots of work has been done on extracting factual information -- the who, what, where, when," explained Claire Cardie, Cornell professor of computer science, who is one of three co-principal investigators for the grant. "We're interested in seeing how we would extract information about opinions."

Cardie is an expert on "information extraction," in which computers scan text to find meaning in natural language. Computer programmers and science fiction fans know that computers are usually very literal and demand that information be presented according to rigid rules. Humans,

on the other hand, are capable of understanding that "Please pass the salt," "May I have the salt," "Hey, is there any salt down there?" and "Yuk, this really needs salt" all mean much the same thing. Cardie's computer programs try to bridge the gap by identifying subjects, objects and other key parts of sentences to determine meaning.

The new research will use machine-learning algorithms to give computers examples of text expressing both fact and opinion and teach them to tell the difference. A simplified example might be to look for phrases like "according to" or "it is believed." Ironically, Cardie said, one of the phrases most likely to indicate opinion is "It is a fact that ..."

The work also will seek to determine the sources of information cited by a writer. "We're making sure that any information is tagged with a confidence. If it's low confidence, it's not useful information," Cardie added.

In addition to the research project, Cardie said, the new UAC has educational goals, seeking to train students to work in information extraction and presenting seminars and workshops for other researchers. The center also will offer summer seminars for women and underrepresented minority undergraduates.

The Department of Homeland Security has established the UACs, Cardie said, partly because it currently lacks enough in-house expertise in natural-language processing. Although the research may conjure fears about invasions of privacy, Cardie says she will be working only with publicly available material, primarily news reports and editorials from English-language newspapers worldwide.

"The techniques would have to be changed considerably to work on documents like e-mails," she noted.

The results, she added, will always include pointers to the original sources, so that when a computer draws some conclusion, human beings will be able to look at the original material and determine whether or not the conclusion was correct.

Source: Cornell University

Citation: Sorting facts and opinions for Homeland Security (2006, September 24) retrieved 9 April 2024 from <https://phys.org/news/2006-09-facts-opinions-homeland.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.