

# The newest AI computing tool: people

June 28 2007

---

A USC Information Sciences Institute researcher is among a growing group of computer scientists learning to solve difficult IT problems of information classification, reliability and meaning by datamining public websites like Digg, del.icio.us and Flickr.

That tool, according to ISI computer scientist Kristina Lerman, is people, human intelligence at work on the social web, the network of blogs, bookmark, photo and video- sharing sites, and other meeting places now involving hundreds of thousands of individuals daily, recording observations and sharing opinions and information.

Lerman shared her recent work with others in the burgeoning new field of social information processing a special AAAI-sponsored symposium on the subject March 26-28 at Stanford.

She says that extracting 'metadata' about transactions -- who is talking to whom, who is listening, how conclusions are reached, and how they spread -- can help researchers answer currently refractory problems about documents: their accuracy and quality, their categorization, the relation of their embedded terminology.

One benefit, according to Lerman, who in addition to her ISI appointment, is a research assistant professor at the Viterbi School of Engineering Department of Computer Science at University of Southern California, is automatic determination of the semantics of content from one kind of metadata: tags.

Tags play a crucial role in a longrunning project called the Semantic Web.

For about a decade, she notes, researchers sought a way to organize data so that someone searching for a specific kind of "check" wouldn't have to weed out unwanted references to chess, symbols, verification procedures, financial documents, political science theories and many more.

Tagging seeks to eliminate ambiguities by affixing 'tags,' computer labels peeling apart the multiple meanings of ordinary language into discreet indicators of meaning, guiding computer searches.

But with natural language being as complex as it is, making sense of tags is not easy. Attempts to manually attack the vocabulary and build in the intricate interconnections that signal different word meanings have proved frustrating.

Lerman hopes she's onto another way. Hundreds of thousands of users are now online, chattering away on all kinds of topics. This volume of directed discourse provides a new way to extracting meaning from tags—statistical models.

The process has been called "folksonomy," a collectively constructed informal classification system. Unlike the traditional approach to the Semantic Web, in which a few knowledge professionals try to agree on a formal classification system which will then be used to annotate data, folksonomy emerges from collective tagging activities of many individuals.

New social websites aimed at sharing information such as del.icio.us and Flickr organically grow ways for site members to access each others holdings. Typically, the members themselves spontaneously create a

tagging system, encouraged by the site architecture.

The tags emerging from such systems, Lerman and collaborators have found, can be turned to broader purposes.

One of Lerman's initial tagging investigations used the photo-sharing site Flickr, analyzing results returned by a request for images of 'beetles,' including some pictures of insects, some pictures of Volkswagens, and a few other entries.

By extracting the tags that Flickr users had described the images with, and applying a mathematical technique called the "Expectation-maximization (EM) algorithm," Lerman found it possible to quite accurately separate pictures of insects from pictures of cars returned by the "beetle" search.

Lerman has gone beyond tagging to using metadata to acquire more and more accurate information about the content of documents in social networking situations.

A Lerman paper now in pre-publication on "Social Information Processing in Social News Aggregation" notes: The rise of the social media sites, such as blogs, wikis, Digg and Flickr among others, underscores the transformation of the Web to a participatory medium in which users are collaboratively creating, evaluating and distributing information.

The innovations introduced by social media have lead to a new paradigm for interacting with information, what we call 'social information processing'.

In the paper, Lerman argues that "by tracking stories over time, that social networks play an important role in document recommendation." In

addition to providing a platform for document recommendation, social Web enables researchers to study collective user behavior quantitatively.

In the same paper, Lerman also presented a mathematical model of how collaborative rating and promotion of stories emerges from the independent decisions made by many users. She found good agreement between predictions of the model and user data gathered from Digg.

In another paper, examining de.licio.us, Lerman and collaborators "describe a probabilistic model of the user annotation process," and then used the model "to automatically find resources relevant to a particular information domain ... with promising results."

Source: University of Southern California

Citation: The newest AI computing tool: people (2007, June 28) retrieved 9 April 2024 from <https://phys.org/news/2007-06-ai-tool-people.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.
---