

measures usage figures of over a hundred digital library repositories used by computers in the university network, which includes about 10,000 computers.

The application consists of three parts. The server part runs in the background, communicates with the network probe, collects and aggregates filtered data, and prepares them for visualization. The Web part takes the data and creates reports, which are available on the Masaryk University library Web site. The third part comprises an analytical desktop application that uses advanced methods of visualization to combine dynamic graphs, tables, forms and statistical graphs. This enables users to see the communication between faculties and digital library servers in a dynamic mind map: faculties and digital library servers are shown as nodes, while edges

between nodes represent their communication. The size of the node illustrates the amount of data transferred to and from the node. The thickness of the edge shows the amount of data transferred during the communication between two nodes. The values are relative to the time period a user selects for analysis. A user can interactively open details about each node and edge and see detailed usage graphs for faculties or digital libraries.

This work is just the beginning. We look forward to analysing usage data over longer periods, seeing trends and anomalies, and thoroughly evaluating the use of digital libraries at Masaryk University. In the future, we also plan to enrich the application with other data sources like the university information system or to include reports from digital library providers.

The application MyLibScope is the result of a broader project, which aims to enhance work with digital libraries both for end users and library administrators. The internal Masaryk University project 'Digital Libraries at Masaryk University' started at the beginning of 2007. MyLibScope was developed and deployed at the end of 2008 and will run in pilot phase from the beginning of 2009. The application uses hardware probes by InveaTech a.s. and is built on software technology by MycroftMind a.s.

Please contact:

Jiří Šmerda

Institute of Computer Science, Masaryk University, Czech Republic

Tel: +420 549497676

E-mail: smerda@ics.muni.cz

Multi-Faceted Rating of Product Reviews

by Stefano Baccianella, Andrea Esuli and Fabrizio Sebastiani

Researchers from ISTI-CNR, Pisa, are working on an automatic rating system for online product reviews based on an analysis of their textual content..

Online product reviews are now available on a variety of Web sites, and are being used by consumers with increasing frequency in order to make purchase decisions between competing products. For example, according to a study performed on TripAdvisor (one of the most popular online review sites for tourism-related activities) of users of the TripAdvisor online booking system, 97.7% are influenced by other travellers' reviews, and of those, 77.9% use the reviews as an aid in choosing the best place to stay.

It is obvious, therefore, that there is a growing market for software tools that can organize product reviews and make them easily accessible to prospective customers. Among the issues that the designers of these tools need to address are: (a) content aggregation, such as pulling together reviews from sources as disparate as newsgroups, blogs and community Web sites; (b) content validation, as in filtering out fake reviews authored by people with vested interests; and (c) content organization, as in automatically ranking competing products in terms of the satisfaction of consumers who have already purchased the product.

We address a problem related to issue (c), namely rating, which involves attributing a numerical score of satisfaction to consumer reviews based on their textual content. This problem arises from the fact that while some online product reviews consist of a textual evaluation of the product and a score expressed on some ordered scale of values, many other reviews only contain a textual part. Such reviews are difficult for an automated system to manage,

especially when a qualitative comparison of them is needed in order to determine whether product x is better than product y, or to identify the best product of the lot. Tools capable of interpreting a text-only product review and scoring it according to how positive it is, are thus of the utmost importance.

Our work looks at the problem of rating a review when the value to be attached must range on an ordinal (ie discrete)



Example of "product" rating.

scale. This scale may be in the form either of an ordered set of numerical values (eg one to five stars), or of an ordered set of non-numerical labels (such as 'poor', 'good', 'very good', 'excellent'). We also focus on multifaceted rating of product reviews, where the review of a product (eg a hotel) must be rated several times according to several orthogonal aspects (eg cleanliness, location etc).

We focus on generating the vectorial representations of the reviews that must be given as input to the learning device used to generate a review rater, rather than on the learning device itself (for which we use an off-the-shelf package). These representations cannot simply consist of the usual 'bag of words' used when classifying texts by topic, since classifying texts by opinion (which is the key content of reviews) requires a much subtler approach. Two expressions such as "A great hotel in a horrible

town!" and "A horrible hotel in a great town!" would receive identical 'bag of words' representations despite expressing opposite opinions.

We have focused on three aspects of the generation of meaningful representations of product reviews: (i) the extraction of complex features based on speech patterns; (ii) making the extracted features more robust through the use of a lexicon of opinion-laden words; and (iii) the selection of discriminating features through techniques explicitly devised for ordinal regression (an issue which until now has received practically no attention in the literature). In order to test the techniques we have developed, we crawled the Web to create a dataset of hotel reviews. The dataset is now available to the research community for experimentation. Several experiments that we have run on it confirm that a combination of these three techniques provides

the best performance on this particular type of data.

The system we have realized could work as a building block for other larger systems that implement more complex functionality. For instance, a Web site containing product reviews whose users only seldom rate their own reviews could use our system to learn from already rated reviews how to rate the others; another Web site containing only unrated product reviews could learn to rate its own reviews, from the rated reviews of some other site.

Link:

<http://nmis.isti.cnr.it/sebastiani/Publications/ECIR09c.pdf>

Please contact:

Fabrizio Sebastiani
ISTI-CNR, Itlay
Tel: +39 050 3152 892
E-mail: fabrizio.sebastiani@isti.cnr.it

More on Mobile Telephones and Our Health

by Harry Rudin

For well over two years, it has been hoped that the international investigation known as the Interphone Study would shed much-needed light on the question of whether mobile phone usage is a threat to our health. While the final results are not yet in, an excellent evaluation of the current status of the study has just been published. The indications are that if there are health dangers, they likely come from intensive, long-term use.

The title of this comprehensive, well-written and easily digestible report is 'Comments on the Interphone Study'. It is available in German, French, Italian and English and was written by Gregor Dürrenberger, Jürg Fröhlich and Heinz-Gregor Wieser in January 2009. The authors are with, respectively, the Swiss Research Foundation on Mobile Communication, the Laboratory for Electromagnetic Fields and Microwave Electronics - ETH Zurich, and the Department of Neurology, University Hospital Zurich, all in Switzerland. The report was published under the auspices of the Research Foundation for Mobile Communication and the Krebsliga Schweiz, a Swiss organization for cancer research and education.

The topic is controversial, delicate and complicated, to the point that the many researchers involved in the Interphone Study have been unable to reach an



Is mobile phone usage is a threat to our health?

overall conclusion. Many individual national studies have been published however, and the cited report describes these results with remarkable clarity. The hope remains that there will eventually be a single report which will combine the many national studies, thus including more individual investigations and so leading to a more robust result. Statistically speaking, individually the national studies examine an

insufficient number of cases to reach a solid conclusion. The reasons for this are also explained in the report.

The goal of the Interphone Study was to examine the potential link between mobile phone use and the risk of developing four different kinds of tumours in the head of the user. The approach used was epidemiological, using case-control studies. People were identified who had one of the four kinds of tumours and an investigation was made into their mobile phone use. The results were compared with those of demographically similar people who had not developed a tumour. The goal was to establish a connection between mobile phone use and the development of these tumours.

Results

Since the individual national studies involved relatively small numbers of people, the certainty of the conclusions,