

Information Retrieval vs Knowledge Retrieval: A Social Network Perspective

Bettina Hoser

Information Services and Electronic Markets,
Institute of Information Engineering and Management,
Department of Economics and Business Engineering,
Universität Karlsruhe (TH)
Germany
hoser@iism.uni-karlsruhe.de

Bettina Hoser
Institute for Information services and management
Department for Information Services and Electronic Markets
Universität Karlsruhe (TH)
Kaiserstrasse 12
D-76128 Karlsruhe
Germany

Information Retrieval vs Knowledge Retrieval: A Social Network Perspective

Abstract. Information Retrieval, especially in connection with the internet, is a well known research field. But as the technologies used for the internet become more and more elaborate, so grows the need to not only find (retrieve) already available information, but to generate knowledge. In this paper the emergence of collective intelligence from information retrieval and social network analysis will be presented. The main point will be how information gained from e.g. websites can be enhanced to become knowledge using methods and models from the research field of Social Network Analysis. Such an approach is very context sensitive, so two examples will be presented.

1 Introduction

When is a trend a trend? When 'the right people' initialize it. This is very well known from the world of fashion. In the world of news, research and technology this may translate to the fact that a trend is a trend when 'relevant' people or websites take up the topic. But how can the relevant people or websites be distinguished from the less relevant? How can 'relevant' be defined? How can one detect really 'relevant' trends? 'Relevant' is always a reflective approach. It is dependent on the circumstances. Thus it is, e.g. in the case of fashion or news, a social context.

As an example for the question discussed here take a high tech company (e.g. mobile phones) or a reinsurance company. For both it is essential that they see trends before the competitors or the possible clients see it. In the case of the high tech company, for example, it is crucial to know what the potential customers are interested in, or which features in the current product are not accepted and why. For the reinsurance company, it is necessary to know which hazards, e.g. in health care, are being discussed, so that the company may prepare its policy accordingly. As an illustration of that point take the discussion on obesity in children and subsequent health problems in adults.

2 Information retrieval

As companies look for ways to find trends as shown above they used to look for example at newspapers. Nowadays the internet with its chat rooms, newsgroups, social networking sites and blogs offers a wide area of information, which had not been accessible before. To gather this information various methods have been devised.

Text analysis is one of the methods often used to extract information from a text source. There is a large body of research literature, see e.g. [FNR03], in the fields of linguistics, information science or classification on diverse ways

to extract keywords, key phrases, etc. from websites and other text sources. In these research fields models have been built to explain how the context sensitive relevance of words, phrases etc. can be defined. Just think about classifications like e.g. the ACM Classification System. Some of these methods lead to lists of possible topics listed by relative relevance according to their usage of phrases in text.

Another approach is to use additional information like keyword or tags to enhance the information retrieved by classifying it. This has grown into the research fields on folksonomies, tagging, semantic web, etc.

At this point though what is known is that these phrases or words are often used. What is not known is who used them. Or to put it precisely, whether the user is a 'relevant' user in the context. This is a question that has been at the heart of the research field of Social Network Analysis.

3 Social Network Analysis

Social Network Analysis (SNA) is a research area that tries to analyze and model actor behavior based on his or her connections or relations to other members of a group. For further reference see [WF99]. An actor is thus seen as restricted or empowered by his or her connection to others. The basis of this structural approach is given by models about group interaction. The first research questions were posed to define roles to actors given a social context. Thus e.g. leadership of a group is such a role. There are also models about the power to manipulate. Thus a person in such a context may be called relevant, or central, if he or she is positioned in such a way in the group's network that all information exchanged between any two actors has to pass through this 'central' actor. He or she can thus manipulate the group.

Thus the question of who is relevant within a group is one of the research questions with SNA. Based on graph theory this can be analyzed by using different so called centrality indices. Some of them are intuitive, like e.g. degree centrality, other are more elaborate like e.g. betweenness centrality or eigenvector centrality. But always the question is: given a clearly defined context, who within a group is relevant, who is not, how are the actors in the group connected and what, if any, predictions can be made for the future development of the group structure.

Thus, this analysis approach can be used to find the 'relevant' people or websites needed to enhance the information found by text retrieval.

4 Knowledge Retrieval

The idea to retrieve knowledge means not only to gather the information available but to enrich it with other information to gain knowledge about a topic. In the case proposed here this means to use results from SNA to enrich the information gathered by text analysis to find whether the topics found by information retrieval are 'really hot topics' because 'relevant people' talk about it, or

whether it is just 'small talk' by 'bystanders'. In a conceptual study [HSGS⁺07] we used such an approach to look for socially enriched information about mobile phones within a newsgroup.

The idea proposed here is based on following information fusion approach: First a text corpus and a group are defined. Then the text corpus is analyzed and the group structure is evaluated. As a last step these two results are combined to gain knowledge. This is just a very crude and short description of the procedure. One major challenge here is to define the group. Depending on the area of interest this can be a very large group or a collection of websites corresponding to a group. Sometimes this may not even be a well defined group. Thus biases can be introduced by choice of actors (or websites). But once the group is defined, there is also the question of the appropriate text analysis method. Questions like scalability and validity have to be answered here. As a last step, the interpretation of the combined results have to be validated before any measures should be taken.

But even with regard to the aforementioned challenges this approach seems to yield deeper insights into topics and trends, since it includes the social component of trends.

5 Outlook

The potential of such an approach is very high. Not only are companies interested in such a kind of knowledge gained from different 'news'-sources, weighted by the social impact, but also the average internet user. If one takes a look at communities of diverse interests such as travel or such necessities as emergencies, it is not only valuable to have information at hand gathered from collective sites, but also to know who gave the information and whether the source can be viewed as 'relevant' in the given context. In the context of emergencies, this may save lives.

References

- [FNR03] J. Franke, G. Nakhaeizadeh, and I. Renz, editors. *Text Mining, Theoretical Aspects and Applications*. Advances in soft computing. Physica-Verlag, Heidelberg, 2003.
- [HSGS⁺07] B. Hoser, J. Schröder, A. Geyer-Schulz, M. Viermetz, and M. Skubacz. Topic trend detection in newsgroups. *Künstliche Intelligenz*, (3), 2007.
- [WF99] S. Wasserman and K. Faust. *Social Network Analysis*, volume 8 of *Structural Analysis in the Social Sciences*. Cambridge University Press, Cambridge, 1 edition, 1999.