

OPEN SEMINAR SERIES

THURSDAY, 19th February 2009
12:00 – 13:00

SEERC Seminar Room
SEERC Bldg

Graph Based Term Weights for Information Retrieval: Text as a Network of Words

By

Dr Christina Lioma

Postdoctoral Fellow

Department of Computer Science

Language Intelligence and Information Retrieval Group

Katholieke Universiteit Leuven

ABSTRACT

The task of an Information Retrieval (IR) system is to retrieve documents from a large repository of data, which are relevant to a user query. This task is addressed by matching documents to queries on a term basis, and by ranking the documents accordingly. A core component of this process is the use of "term weights", which are weights representing how much a term contributes to the meaning of the text where it occurs.

Typical term weights are computed using lexical frequency statistics, i.e. word counts. This talk will present a different type of term weights, namely "graph based term weights". The computation of such weights involves modeling text as a graph, where vertices denote terms, and edges denote co-occurrence and grammatical relations between terms.

Modeling text as a graph is an interesting alternative to modeling text as a bag of words, and allows to compute term weights that contain statistical or linguistic relations as an integral part of their computation. Experimental evaluation confirms the usability of graph based term weights for IR systems.

Christina Lioma is a postdoctoral fellow at Katholieke Universiteit Leuven (Belgium), Department of Computer Science, Language Intelligence and Information Retrieval group. She holds a PhD in Computer Science from the University of Glasgow, Information Retrieval group, an MSc in Natural Language Processing from the University of Manchester, and an MA Hons in Languages and Linguistics from the University of Glasgow. Her research focuses on information retrieval, textual natural language processing, and quantitative linguistics.

<http://www.cs.kuleuven.be/~christin/>

The seminar series is open to all members of *staff* and *students* of CITY and to any *externals* that wish to attend.