

Keyword Extraction in Information Retrieval

Harold Short (*harold.short@kcl.ac.uk*)

King's College London

Marilyn Deegan (*marilyn.deegan@kcl.ac.uk*)

King's College London

Laszlo Hunyadi (*hunyadi@ling.arts.klte.hu*)

University of Debrecen

Paul Baker (*p.baker@lancaster.ac.uk*)

Lancaster University

Dawn Archer (*d.archer@lancaster.ac.uk*)

University of Central Lancashire

Tony McEnery (*a.mcenery@lancaster.ac.uk*)

Lancaster University

Session chair: Harold Short.

Keyword extraction: an Overview

Laszlo Hunyadi

With the ever increasing amount of information made available in all areas of life, including economy, science, education and culture, there is an imperative need to retrieve, elaborate and present this information in the most optimal way. Since the bulk of information is textual, information retrieval is mainly concerned with texts. This talk will present an outline of the approaches, principles and techniques used in textual data retrieval based on keyword extraction. There will be an analysis of the capabilities of various approaches showing their appropriate uses.

Keyword extraction in the project Forced Migration Online

Marilyn Deegan

Forced Migration Online (<http://www.forcedmigration.org/>) provides instant access to a wide variety of online resources dealing with the situation of forced migrants worldwide. Being one of the most comprehensive textual resources dealing with humanitarian issues across a large number

of countries, it faces the challenges of multilingualism, multiculturalism and the essential requirement to be up-to-date and informative. That is why the organisation and retrieval of textual information as well as operability have high priority in the design and functioning of the system. This talk presents the essentials of this online resource, including its pioneering beginnings and view of future development.

Querying Keywords: Questions of difference, frequency and sense

Paul Baker

This paper examines issues to do with interpreting keyword lists (Scott 1999), such as over-attending to lexical differences whilst ignoring differences in word usage and/or similarities between texts. Using a variety of techniques (e.g. analysis of key clusters or annotated data), I show how researchers can use keyword analyses to obtain a more accurate picture of the distinctive features of their texts or corpora.

Love - a familiar or a devil? An exploration of key domains in Shakespeare's Comedies and Tragedies

Dawn Archer, Jonathan Culpeper, Paul Rayson

Love is a common theme in Shakespeare's works. In this paper, we show how the *UCREL Semantic Annotation Scheme* (henceforth *USAS*), a software program for automatic dictionary-based content analysis, can help us to explore the semantic field of 'love' within a selection of Shakespeare's plays. Specifically, we will explore 3 love-tragedies (*Othello*, *Antony and Cleopatra*, and *Romeo and Juliet*) and 3 love-comedies (*A Midsummer Night's Dream*, *The Two Gentlemen of Verona* and *As You Like It*) to determine differences in their (re)presentation of 'love'. We will also discuss how the semantic field of 'love' co-occurs with different domains in the plays, and assess the implications this has on our understanding of 'love' as a concept. This research builds on (i) Jonathan Culpeper's work on keywords in Shakespeare, using *Wordsmith* (Culpeper 2002), (ii) Paul Rayson's comparisons of key word and key domain analysis (Rayson 2003), and (iii) Dawn Archer and Paul Rayson's work on the identification of key domains in refugee literature, using *USAS* (Archer and Rayson forthcoming).

Key words and the analysis of discourses in historical contexts

Tony McEnery

This paper examines the use of keywords to approach the discourse of moral panic evident in the writings of the Society for the Reformation of Manners in late seventeenth/early

eighteenth century England. The keyword approach, I will argue allows one to populate a model of moral panic discourse, while simultaneously showing how, in that historical context, links were forged between concepts which, while unlinked then, have become naturalised as being linked in modern English. By showing how keywords relate to discourse, and ultimately to a process whereby meanings and objects become linked, the paper will argue that keywords are important tools for the historical linguist in studying the shifting patterns of word association in language.

Bibliography

Archer, D., and P. Rayson. "Using the UCREL automated semantic analysis system to investigate differing concerns in refugee literature." *The Keyword Project: Unlocking Content Through Computational Linguistics*. Ed. M. Deegan, L. Hunyadi and H. Short. Office for Humanities Communication Publications, Forthcoming.

Culpeper, J. "Computers, language and characterisation: An analysis of six characters in *Romeo and Juliet*." *Papers from the ASLA symposium, Conversation in life and literature*. Ed. Ulla Melander Marttala, Carin Ostman and Merja Kyto. Uppsala: Association Suedoise de Linguistique Appliquee, 2002. 11-30.

Rayson, P. "Matrix: A statistical method and software tool for linguistic analysis through corpus comparison." Ph.D. thesis, Lancaster University.