Text Mining: beyond the CAQDAS tools?

Davy Weissenbacher, Brian Rea, Sophia Ananiadou

National Centre for Text Mining
Manchester Interdisciplinary Biocentre
University of Mancester
131 Princess Street, M1 7DN
{firstname.lastname}@manchester.ac.uk

Keywords: CAQDAS tools, Text mining, textual document

Abstract: The interpretation and analysis of qualitative data in literature is an important yet time consuming task in current research practice. Specialised tools, such as those collected around the CAQDAS networking project[1], have been developed to assist this process through the common principles of managing heterogeneous documents and annotating the content. Whilst this software can improve the efficiency of the researcher, the process is still mostly manual. Therefore the analysis of larger document collections, which are often required for constructing theories, remains an activity taking a great deal of time and effort from the researcher.

There exists a potential solution for automating part of this process for textual documents through a combination of text mining techniques. Using this technology the user can make better use of the resources at hand through clustering and visualisation of the collection of documents and to ease the burden of large-scale annotation by learning from the salient features both automatically detected and from samples of user annotations.

In this article we examine the strategies currently offered in a selection of CAQDAS tools. We show that some tools already extend their functionalities with the addition of text mining, ranging from the statistical content analysis provided by *QDA miner[2]* to a system supporting automatic learning of user annotations with *Qualrus[3]*. Following a discussion of the strengths and limitations of these tools we expand on the methodology proposed above and examine the current implementation in the system realized in the ASSIST[4] project.

---

[1]   Computer Assisted Qualitative Data Analysis Software - http://caqdas.soc.surrey.ac.uk/

[2]   QDA Miner - http://www.provalisresearch.com/QDAMiner/QDAMinerDesc.html

[3]   Qualrus - http://www.ideaworks.com/qualrus/index.html

References

S. Ananiadou, R. Procter, B. Rea, Y. Sasaki, and J. Thomas (2007) *Supporting Systematic Reviews using Text Mining*, in 3rd International Conference on e-Social Science, Ann Arbor.

K. Frantzi, S. Ananiadou and H. Mima (2000) *Automatic recognition of multi-word terms: the C-value/NC-value method*, International journal on Digital Libraries 3/2 117-132.

Y. Sasaki, B. Rea and S. Ananiadou (2007) *Multi-topic aspects in clinical text classification*, in IEEE BIBM 2007.

A. Lewins and C. Silver (2006) *Choosing a CAQDAS Software Package 5ᵗʰ Edition*, a working paper.

---

[4]   Further information on this JISC funded project can be found at http://www.nactem.ac.uk/assist/