

Are the CAQDAS and the Text Mining software competitors?

The Computer-assisted Qualitative Data Analysis Software (CAQDAS) is a generic term to denote a set of qualitative analysis software used in different areas of the humanities. Text Mining (TM) refers to software designed to automatically extract information from human written documents.

Concretely, TM and CAQDAS software both annotate documents and can be regarded as competitors. We claim that not only they are not competitors, but they are complementary software.

The CAQDAS software has been created to facilitate a human analysis of a small corpus (hundreds of documents). TM software performed an automatic annotation of a huge corpus (million of documents). Annotations added to the original documents by the software belong to complementary levels: the former add semantic and pragmatic annotations, the latter lexical, syntactic and simple semantic annotations.

CAQDAS software tools already integrate basic TM techniques to extend their research interface to words and documents using a lemmatizer, thesaurus and clustering algorithm. The TM software can also benefit from the CAQDAS software using the human annotations that have supplementary pieces of information for the automatic learning algorithms.

Examining the current implementation of the TM system for a frame analysis of mass-media documents during the ASSIST project, we will show how CAQDAS software can be integrated to extend and optimize the automatic learning algorithms.