Top free Text processing software:

1. QDA miner lite:

It is a free computer assisted qualitative analysis software from Provalis research. It can be used for analysis of textual data as well as still images. It offers basic CAQDAS features like importation of documents in plain text, RTF, HTML, PDF as well as data stored in MS Excel, Access, CSV, tab delimited files. It provides fast Boolean text search tool for retrieving and coding text segments and code frequency analysis with bar chart, pie chart and tag clouds.

1. GATE:

General Architecture for Text Engineering is an open source software capable of solving almost any text processing problem. It is a defined and repeatable process for creating robust and maintainable text processing workflows. In active use for all sorts of language processing tasks and applications, including: voice of the customer; cancer research; drug research; decision support; recruitment; web mining; information extraction; semantic annotation. (Top 20 free software for Text analysis, Text mining, Text analytics - Predictive analytics today, 2014)

GATE is an ideal starting point for student projects on language analysis, as it comes with a set of Information Extraction modules that can be used as a base, and a significant number of PhD students have used GATE in their research.

1. TAMS Analyzer:

TAMS stands for Text Analysis Markup System. It is a convention for identifying themes in texts (web pages, interviews, field notes). It was designed for use in ethnographic and discourse research.

TAMS Analyzer is a program that works with TAMS to let you assign ethnographic codes to passages of a text just by selecting the relevant text and double clicking the name of the code on a list. It then allows you to extract, analyze, and save coded information. TAMS Analyzer is open source; it is released under GPL v2. The Macintosh version of the program also includes full support for transcription (back space, insert time code, jump to time code, etc.) when working with data on sound files.

1. Carrot2:

Carrot2 does text and search results clustering framework. Carrot2 is an Open Source Search Results Clustering Engine. It can automatically organize small collections of documents (search results but not only) into thematic categories.

1. CAT:

CAT is a free service of the Qualitative Data Analysis Program (QDAP). CAT allows you to efficiently code raw text data sets, annotate coding with shared memos and allows you to export coding in RTF, CSV or XML format.

1. KH Coder:

KH Coder is a free software for quantitative content analysis and text data mining. KH coder can analyze Japanese, German, English, French, Italian, Portuguese, and Spanish texts. The Input raw texts, can utilize searching and statistical analysis functionalities like KWIC, collocation statistics, co-occurrence networks, self-organizing maps, multidimensional scaling, cluster analysis and correspondence analysis.

1. Tm- text mining package:

Tm provides a framework for text mining applications within R. The tm package offers functionality for managing text documents, abstracts the process of document manipulation and eases the usage of heterogeneous text formats in R. The package has integrated database back-end support to minimize memory demands. An advanced Meta data management is implemented for collections of text documents to alleviate the usage of large and with Meta data enriched document sets. The package provides native support for reading in several classic file formats (e.g. plain text, PDFs, or XML files). There is also a plug-in mechanism to handle additional file formats. There is also a plug-in mechanism to handle additional file formats. The data structures and algorithms can be extended to fit custom demands

1. Rapidminer:

Rapidminer is an integrated environment for machine learning, data mining, text mining, predictive analytics and business analytics. Rapidminer text extension adds all operators necessary for statistical text analysis. Texts from different data sources can be loaded and, can be transformed by different filtering techniques, to analyze data.

The Rapidminer studio Starter Edition is an excellent software tool for predictive text mining and sentiment analysis. It is free and provides all the major algorithms used for different types of text mining techniques on your text data.

1. Unstructured Information Management Architecture(UIMA):

Unstructured Information Management Architecture is a component framework to analyze unstructured content such as text, audio or video. This is originally developed by IBM.

UIMA enables applications to be decomposed into components, for example "language identification" => "language specific segmentation" => "sentence boundary detection" => "entity detection (person/place names etc.)". Each component implements interfaces defined by the framework and provides self-describing metadata via XML descriptor files. The framework manages these components and the data flow between them. Components are written in Java or C++; the data that flows between components is designed for efficient mapping between these languages.

1. OpenNLP:

The Apache OpenNLP library is a machine learning based toolkit for the processing of natural language text.

It supports the most common NLP tasks, such as tokenization, sentence segmentation, part-of-speech tagging, named entity extraction, chunking, parsing, and coreference resolution. These tasks are usually required to build more advanced text processing services. OpenNLP also includes maximum entropy and perceptron based machine learning.

1. Apache Mahout:

Apache Mahout is a project of apache software foundation with the objective of creating scalable machine learning algorithms that are free to use under the Apache license. The three major components of Mahout are an environment for building scalable algorithms, many new Scala + Spark (H2O in progress) algorithms, and Mahout's mature Hadoop MapReduce algorithms.

References:

* (Top 20 free software for Text analysis, Text mining, Text analytics - Predictive analytics today, 2014)
* (Carrot 2 Open source framework for building search clustering engines, n.d.)
* (General Architecture for Text Engineering, n.d.)
* (QDA- Miner Lite)
* (tamsys.sourceforge.net)
* (Coding Analysis Toolkit)
* (KH- Coder Predictive Analytics Today)
* (tm- text mining package)
* (Apache.org)
* (Apache OpenNLP)
* (Apache Mahout- Scalable machine learning and data mining)