

Announcement

Quantitative Index Text Analyser (QUITA)

Miroslav Kubát, Vladimír Maltach, Radek Čech
(Palacký University, Olomouc)

New software for a quantitative text analysis has been developed at Palacký University in Olomouc, the Czech Republic. *Quantitative Index Text Analyser* (QUITA) covers the most common indicators, especially those connected with frequency structure of a text. In addition to computing results of the indicators, QUITA provides also statistical testing and graphical visualization of obtained data.

QUITA is a versatile tool with many uses designed for researchers from various disciplines (linguistics, criticism, history, sociology, psychology, politics, biology, etc.). The program enables basic text processing functions like creating word lists, text lemmatizing or creating n-grams. The program also provides more advanced tools such as a random text creator or a binary file translator. However, the main part of the software is an indicator computing. Although the authors focused mainly on the indicators connected to frequency structure of a text (e.g. *h*-point, entropy, repeat rate, adjusted modulus, Gini's coefficient, lambda), there are also several other characteristics such as thematic concentration, activity & descriptivity or writer's view.

The main purpose of QUITA is to provide user-friendly tool of quantitative text analysis for researchers (especially from the humanities) without deeper knowledge of quantitative linguistics, statistics and programming. Apart from generating results, QUITA also enables a simple statistical comparison and creating charts. There is no need to use any additional software such as spreadsheet applications or special statistical programs. In sum, QUITA is the program that combines all important parts of any quantitative research: obtaining results, statistical testing and graphical visualization.

In order to compare texts for authorship attribution, genre analysis or another purpose, the differences between obtained resulting values of several indicators can be statistically tested. QUITA provides not only statistical testing among particular texts but also among groups of texts. For creating graphs of obtained data, there is a special tool "Chart Wizard" which offers wide range of chart types and editing options. All results can be copied via clipboard or saved directly as CSV file. The charts can be saved as image files.

QUITA is a tool with wide range of application, from stylometry to DNA analysis. Although almost all indicators in the software were proposed as features for common linguistic research (e.g. authorship attribution, genre or thematic analysis), possibilities are practically endless. Biologists can use one of available tokenizers (DNA Triplet Tokenizer, DNA Nucleotide Tokenizer) to handle with DNA as a text and apply the indicators, for instance. There is also an option to use different units other than words or lemmas such as characters, n-grams, etc. It should be noted that the software is designed as multilingual tool; QUITA therefore works with almost all

scripts and includes several tokenizers and lemmatizers. Nevertheless, especially the number of lemmatizers is still limited but it should be significantly extended in a next version of the software.

Since QUITA aims to help as many researchers as possible, the program will be distributed as freeware. Thus everybody can use QUITA without any restrictions. The software can be downloaded on the website <http://oltk.upol.cz/software>.

The software was developed as a student project at the Department of General Linguistics at Palacký University in Olomouc, the Czech Republic. The team consists of two students (Vladimír Matlach, Miroslav Kubát) and their supervisor Radek Čech. The indicators included in QUITA were mostly selected in accordance with following books: *Word frequency studies* (Popescu et al. 2009), *Aspects of Word Frequencies* (Popescu et al. 2009) and *Metody kvantitativní analýzy (nejen) básnických textů* (Čech et al. 2013).

Acknowledgement

QUITA (Quantitative Index Text Analyser) was supported by the student project IGA (no. FF_2013_031) of the Palacký University, Olomouc, Czech Republic.

References

- Čech, R., Popescu, I. I., Altmann, G. (2013). *Metody kvantitativní analýzy (nejen) básnických textů*. Olomouc: Univerzita Palackého v Olomouci. (in press)
- Popescu, I.-I., Altmann, G., Grzybek, P., Jayaram, B.D., Köhler, R., Krupa, V., Mačutek, J., Pustet, R., Uhlířová, L., Vidya, M.N. (2009). *Word frequency studies*. Berlin-New York: Mouton de Gruyter.
- Popescu, I.-I., Mačutek, J., Altmann, G. (2009). *Aspects of word frequencies*. Lüdenscheid: RAM-Verlag.