Studies in Quantitative Linguistics

Editors

Fengxiang Fan (fanfengxiang@yahoo.com)
Emmerich Kelih (emmerich.kelih@uni-graz.at)
Reinhard Köhler (koehler@uni-trier.de)
Ján Mačutek (jmacutek@yahoo.com)
Eric S. Wheeler (wheeler@ericwheeler.ca)


ISBN: 978-3-942303-28-6

© Copyright 2014 by RAM-Verlag, D-58515 Lüdenscheid, Germany

RAM-Verlag
Stüttinghauser Ringstr. 44
D-58515 Lüdenscheid
Germany
RAM-Verlag@t-online.de
http://ram-verlag.de
Contents

1 Introduction .............................................................................................................................................. 1
2 System requirements ................................................................................................................................. 3
3 Install and Starting the Application ......................................................................................................... 4
4 Creating a New Project ............................................................................................................................. 5
5 Creating and loading texts ......................................................................................................................... 7
5.1 Supported files and encodings .............................................................................................................. 8
6 Indicators to compute ............................................................................................................................... 9
6.1 Frequency Structure Indicators ........................................................................................................... 10
  6.1.1 Type-Token Ratio (TTR) .................................................................................................................. 10
  6.1.2 h-point (h) ....................................................................................................................................... 11
  6.1.3 $R_1$ .................................................................................................................................................. 14
  6.1.4 Repeat Rate (RR) ............................................................................................................................ 16
  6.1.5 Relative Repeat Rate of McIntosh (RR$_{mc}$) ................................................................................. 19
  6.1.6 Hapax Legomena Percentage (HL) ................................................................................................. 20
  6.1.7 Lambda ($\Lambda$) ............................................................................................................................ 21
  6.1.8 Gini Coefficient ($G$) ...................................................................................................................... 30
  6.1.9 $R_4$ .................................................................................................................................................. 34
  6.1.10 Curve length ($L$) ........................................................................................................................ 35
  6.1.11 Curve length R Indicator ($R$) ...................................................................................................... 36
  6.1.12 Entropy ($H$) .................................................................................................................................. 38
  6.1.13 Adjusted Modulus ($A$) ............................................................................................................... 40
6.2 Miscellaneous indicators ....................................................................................................................... 42
  6.2.1 Verb Distances (VD) ....................................................................................................................... 42
  6.2.2 Activity ($Q$) & Descriptivity ($D$) .................................................................................................. 43
  6.2.3 Writer’s View ($a$) ......................................................................................................................... 44
  6.2.4 Average Tokens length (ATL) ....................................................................................................... 47
1 Introduction

Quantitative experimental methods have been increasingly used in the humanities in recent years. We can hardly imagine the disciplines of social science, such as psychology, sociology or economics, without a quantitative approach. On the other hand, the majority of linguists, historians and especially literary critics are still refusing to use quantitative methods. One of the reasons is the fact that those researchers consider quantitative methods, and especially statistical methods, too difficult to apply to their field. QUITA (Quantitative Indicator Text Analyzer) is a tool which aims to help all people who try to analyse texts by quantitative methods.

QUITA is a program to enable researchers from various disciplines (linguistics, criticism, history, sociology, psychology, politics, biology, etc.) to analyse texts using quantitative methods. There are many indicators which measure various characteristics of text. Although the authors of QUITA focused mainly on indicators connected to the frequency structure of a text, there are also functions for several other characteristics. Since QUITA is designed especially for researchers outside quantitative linguistics, it includes functions for the most basic and common indicators.

Given that the main purpose of QUITA is to provide a user-friendly tool of quantitative text analysis for researchers without a deeper knowledge of quantitative linguistics, statistics or programming, QUITA also provides simple statistical comparisons and the ability to create charts. There is no need to use any additional software such as spreadsheet applications or special statistical programs. QUITA is therefore the program that combines all the essential parts of any quantitative research effort: obtaining results, statistical testing and graphical visualization.

The QUITA manual is written with step-by-step instructions. All tools are concisely described and accompanied by screen-shots. Every indicator is briefly presented (complete with references), and mathematically defined. There are also examples of computation and statistical comparison.

Although the manual provides users with all the essential information about QUITA, it was not possible to cover most topics in deeper detail. For this purpose, we highly recommend the book *Word frequency studies* (Popescu et al. 2009) which is a comprehensive overview about quantitative analysis using indicators based on the frequency structure of a text. The book *Aspects of Word Frequencies* (Popescu et al. 2009) is also well worth reading. Detailed examples of computing most indicators used in QUITA can be found in *Metody kvantitativní analýzy (nejen) básnických textů* (Čech et al. 2014).

Since we aim to help as many researchers as possible, QUITA is distributed as freeware. Thus anyone can use QUITA without any restrictions. The latest version of the software is available on the website [https://code.google.com/p/oltk/](https://code.google.com/p/oltk/). In published work, acknowledgement of QUITA would be appropriate and appreciated.
2 System requirements

System requirements: NET Framework 3.5
Optional system requirements: Python, Perl

NOTE:
Users are automatically notified that it is necessary to install the system requirements with the download links.