
The Signature Stylometric System - Estilometria

Autor:

Data de publicació: 02-04-2020

Welcome to the home page of Signature, a program designed to facilitate "stylometric" analysis and comparison of texts, with a particular emphasis on author identification. The collage below on the right illustrates the sorts of task for which Signature can be used: comparing the styles of Jane Austen and other novelists; examining the "authorial signature" of the plays written by (or controversially attributed to) Shakespeare; establishing the provenance of ancient manuscripts such as the shared books of Aristotle's Ethics; identifying the author of the unattributed Federalist Papers; and investigating the relationships between Biblical scriptures (e.g. Did "Luke" write Acts? Did Paul write Hebrews?).

A User-Friendly System for Textual Analysis

Welcome to the home page of Signature, a program designed to facilitate "stylometric" analysis and comparison of texts, with a particular emphasis on author identification. The collage below on the right illustrates the sorts of task for which Signature can be used: comparing the styles of Jane Austen and other novelists; examining the "authorial signature" of the plays written by (or controversially attributed to) Shakespeare; establishing the provenance of ancient manuscripts such as the shared books of Aristotle's Ethics; identifying the author of the unattributed Federalist Papers; and investigating the relationships between Biblical scriptures (e.g. Did "Luke" write Acts? Did Paul write Hebrews?).

Register Your Interest in Signature 2.00

At present (Summer 2013), Signature has been undergoing the most important enhancement since its initial development, which is now very close to completion (testing is in hand, and documentation is 95% completed). Version 2.00 will include a wide range of new facilities, including:

More powerful file-handling and filtering tools

Ability to specify relevant alphabets and punctuation etc. for different languages/genres

Wordlist facilities extended to accommodate phrases of specified length(s)

Similar facilities for bigrams/trigrams etc.

Choice of keyness measures for key word/phrase identification

Fully automatic creation of frequent word/phrase lists

Automated monitoring of previously specified words

Powerful concordancer, enabling also punctuation and proximity searches etc.

Principal Component Analysis, applicable to all data types

Burrows' Delta analysis, applicable to all data types

Multiple chi-square analysis, applicable to all data types

Main parameters of all facilities easily configurable

Comprehensive help and theoretical documentation

Investigation is also under way to test the feasibility of incorporating grammatical analysis into the concordancer, so as to enable grammar-informed searching etc. If this proves feasible, the concordancer will also be further integrated with the graphing and data analysis facilities.

It may be some time before Signature 2.00 is fully tested and published here. In the meantime, if you are interested in acquiring it, please register your interest, so that you can be kept informed of progress and provided with the software at the first available opportunity. You might also be invited (on a purely optional basis, of course) to beta-test the software, assistance with which would be much appreciated.

Download Signature 1.0

This program is freeware for educational use, but please respect the copyright, and ensure that if you pass it on you do so without charge, make clear its authorship, and leave all documentation intact. The program is provided in two forms, first as a standard ZIP archive, and then as a self-extracting ZIP file. In both cases it is packaged together with the Federalist papers, collated by known author, to serve as sample texts for getting started:

Signature 1.0 (with the Federalist Papers) as a standard ZIP archive
Signature 1.0 (with the Federalist Papers) as a self-extracting ZIP file

This is the first publicly available version, but please note that it was at a development stage with a number of important features still to be added and documentation incomplete (e.g. with no online help)

Improvements planned include:

A comprehensive online Help file, giving full explanations of all the system's facilities.
Considerable enhancement of the text filtering mechanisms, to enable the system to deal more intelligently with common textual problems (e.g. those often arising from Web documents or line break variations) and to take advantage of standard markup (e.g. XML/TEI Lite).
Adaptation to non-standard alphabets (e.g. for transliterated Greek) and punctuation (e.g. for Biblical "verses").
Incorporation of Unicode, to enable texts to be processed and displayed appropriately in a wide variety of languages.
Development of the text display facility, to enable further investigation of interesting results unearthed by the analysis.
Addition of concordancing and phrase recognition, as a development of the existing word search facility.
Further statistical operations, including correlation and clustering with appropriate graphical output.

Using the System

Having downloaded the ZIP archive, extract it into an appropriate directory (e.g. "C:Signature") and start the system by running the file "Signature.exe".

Documentation

A PowerPoint presentation is provided in the package, to give a straightforward introduction to the ideas of stylometric analysis and the Signature system in a manner suitable for private study, or a taught course on literary computing. Use PowerPoint to print out handouts (six slides per page) for a useful quick-reference guide:

PowerPoint presentation: Introduction to Textual Analysis using Signature.

Full documentation will in due course be provided in a comprehensive Help file, which is currently in preparation.

Prepared Textual Resources

Although Signature can operate on standard text and HTML files, it is often desirable to prepare these for use appropriately (e.g. by enclosing metadata in "<...>" tag brackets, so as to exclude it from the analysis). This particularly applies to files from the Gutenberg Project, which are otherwise extremely useful for the purpose, but which have extensive front/back matter that needs to be marked out if it is not to distort the stylometric results. The following files contain small archives of pre-prepared files, most of them deriving from the Gutenberg archives:

Novels of Jane Austen, the Brontes, Dickens, and George Eliot, as a standard ZIP archive
Novels of Jane Austen, the Brontes, Dickens, and George Eliot, as a self-extracting ZIP file
Plays of Shakespeare, together with The Two Noble Kinsmen and Edward III (which are of disputed authorship), as a standard ZIP archive
Plays of Shakespeare, together with The Two Noble Kinsmen and Edward III (which are of disputed authorship), as a self-extracting ZIP file
All the books of the Greek New Testament, transliterated into the English alphabet, as a standard ZIP archive
All the books of the Greek New Testament, transliterated into the English alphabet, as a self-extracting ZIP archive