



TextQuest/ReFo manual

October 2024

SSC

Social Science Consulting

<http://www.social-science-consulting.com>

e-mail: info@social-science-consulting.com

Dr. Harald Klein

Lutherstr. 2

49082 Osnabrück

Germany

Tel: +49 541 5979 2545

Contents

1. Introduction	5
1.1 Base version	6
1.2 Standard version	6
1.3 Premium version	6
1.4 Professional version	7
1.5 Combination of available readability formulas	7
2. Working with TEXTQUEST/REFO	9
2.1 Preparing the text	9
2.2 Language selection	9
2.3 Text genre selection	10
2.4 Readership selection	11
3. Inspecting the results	13
4. Vocabulary module	17

1. Introduction

This document describes all versions of TextQuest/Refo, where ReFo is an abbreviation for readability formula. This software is available for MS-Windows and for Apple MacOS. All versions are available for different license times and different users. A text is analysed, the results of an analysis are values of different readability formulas.

In the last 100 years a lot of readability formulas have been developed. All of them are based on a cloze procedure, the results of this analysis is a mathematical formula based on regression analysis. In most formulas the length of sentences and the length of words are the most important variables that explain the readability of a text. Readability formulas were developed for different languages, different text genres, and different readers – therefore you have to know and select the correct formula. TextQuest/ReFo knows each combination and selects the appropriate formula(s) for your needs. However, it may happen that there is no formula for a selected combination of language, text genre, and readership. In total 91 readability formulas are implemented.

The results are dependent on the readability formula(s) and are computed for the whole text and for all its paragraphs. So you can see which paragraphs are better or worse than other ones. The results can be:

- an index value between 0 and 100 – the higher the value, the better the readability of the text. A verbal description is also given so you can interpret the value.
- reading age: the age of the recipient must have to be able to understand the text. This is connected with the school system of the USA.
- reading class: the school class of the reader must have to be able to understand the text. This is connected with the school system of the USA.

TextQuest/ReFo supports:

- languages: English, French, German, Spanish, Italian, Dutch, Afrikaans, Turkish, Polish, Danish, Norwegian, Icelandic, Swedish, Czech, Slovak, Slovenian.
- text genres: unspecific, fictional, non-fictional, technical manuals, children's books, school books, political texts, package inserts of medicine

- readership: unspecific, children, young people, adults, foreign learners, soldiers, patients.

Not all languages, text genres and readers are supported by each version. So it is possible that you only pay for a license that includes what you need.

1.1 Base version

The base versions does not cover all languages, all text genres, or all readerships. The following features are supported:

- languages: English, German
- text genres: unspecific, technical, fictional, non-fictional, children's books, school books, political texts
- readership: unspecific, children up to 10 years, teenagers from 11 to 18 years, soldiers

1.2 Standard version

This version covers all languages, but it does not include the vocabulary module like the professional version.

- languages: English, German, French, Spanish, Italian, Dutch, Flemish, Africaans, Danish, Icelandic, Norwegian, Polish, Swedish, Turkish
- Text genres: unspecific, technical, fictional, non-fictional, children's books, school books, political texts, package inserts of medicines
- Readership: unspecific, children up to 10 years, young people from 11 to 18 years, soldiers, learners of foreign languages

1.3 Premium version

This version doesn't have restriction in languages, text genres, or recipients. Diese Version deckt alle Sprachen ab, enthält aber nicht das Wortschatzmodul wie die Professionalversion.

- languages: English, German, French, Spanish, Italian, Dutch, Flemish, Africaans, Danish, Finnish, Icelandic, Norwegian, Polish, Portugese, Czechm Slovak, Kroatian, Slovenian, Swedish, Turkish

- Text genres: unspecific, technical, fictional, non-fictional, children’s books, school books, political texts, package inserts of medicines
- Readership: unspecific, children up to 10 years, young people from 11 to 18 years, soldiers, learners of foreign languages

1.4 Professional version

Professional: like the standard version, but with an additional vocabulary module. This shows word lists, all combinations of words depending on the number of words, or all permutations within a sentence (which words occur together – co-occurrences).

1.5 Combination of available readability formulas

One feature of itx is that it selects the appropriate formula(s) for many combinations of language, text genre, and readership. However, readability formulas were not developed for every combination. At least one formula that was developed for unspecific text genres and unspecific readerships is available for all languages. Specific formulas are available for the following combinations:

language	combination
Portugese	children older than 10 years
Dutch	non-fictional texts and children’s books for children under10 years (2) all texts and fictional texts for children older than 10 years (2)
Polish	non-fictional texts
Turkish	school books for children under 10 years (2)
English	prose for soldiers (2) technical texts for soldiers (5) prose for children older than 10 years (2) school books for children older than 10 years (3) children’s books for children older younger than 10 years (3) school books for children younger than 10 years (2) school books for all ages
German	political texts for adults all texts for children younger 10 years fictional texts for children older than 10 years (2) non-fictional and fictional texts for children younger than 10 years (8)

Patient leaflets information are available for the following languages: Danish, English, German, Finnish, French, Icelandic, Italian, Norwegian, Slovene, Spanish, and Swedish.

2. Working with TEXTQUEST/REFO

After the start you can see the big editor window. You can paste your text into this window. If your text is in a file, use the file menu to import the text.

2.1 Preparing the text

In most cases you just cut and paste the text into the edit window. If you have HTML, PHP or Javas(script) code in the text, you have to delete it before you start an analysis.

A readability analysis consists of counting grammatical sentences, words and syllables, TEXTQUEST/REFO separates the whole text into grammatical sentences and counts the words and the syllables. You can split the text into paragraphs – just separate paragraphs with a blank line. If you have a heading (with no .?! at the end), make it a paragraph by inserting a blank line.

A lot of abbreviations that use dots are known to TEXTQUEST/REFO. If this is not the case and a new sentence starts after an abbreviation, just remove the dot. It is very unlikely that you will ever have to do this, because a lot of abbreviations – language dependent – are stored in internal lists and recognised not be the end of a sentence but an abbreviation.

Especially in patient leaflets there are enumerations that can produce very long sentences and bias the results. Therefore you can start each enumeration part with a - (minus symbol) at the start of each line that contains a enumeration element.

2.2 Language selection

In the middle of the screen you can select the language, please note that you cannot select all languages in the base version. Abbreviations are language specific, a few thousand are implemented in TEXTQUEST/REFO. In the vocabulary module of the professional version also the correct sort order of the selected language is used, e.g. in German the umlaut ä is treated as its

base vocal a, whereas in Danish ä is sorted behind the z.

Language: English
Genre: All Reader: All
Start Analysis **Clear Results**

- Czech
- Danish
- German
- English
- Spanish
- Finnish
- French
- Croatian
- Icelandic
- Italian
- Dutch, Flamish, Afrikaans
- Norwegian
- Polish
- Portuguese
- Slovak
- Slovene
- Swedish
- Turkish

2.3 Text genre selection

Language: English Genre: All Reader: All
Start Analysis **Clear Results**

- All
- Technical texts
- Fictional texts
- Non fictional texts
- Childrens books
- School books
- Political texts
- Patient leaflet information

The second option is to select the text genre, all means unspecific – applies for all texts. Patient leaflets cannot be analysed with the base version. Many readability formulas were developed for special text genres, the results cannot be used for other text genres. Sometimes also a combination of text genre and readership apply, e.g. fictional texts and children’s books.

2.4 Readership selection

The last option is to select a specific readership, or if that is not the case you select *all*. Some readability formulas are a combination of text genre and readership, TEXTQUEST/REFO selects the appropriate formula(s). There are no more options available.



Language: English Genre: All Reader: All

Start Analysis **Clear Results**

- All
- Children up to 10
- Children from 11 up to 18
- Soldiers
- Foreign language learners

After selecting language, text genre, and readership it may happen that there is no readability formula for that combination. In most cases you select all for reader ship and/or text genre. Now you can press the Analysis button to start the analysis. TEXTQUEST/REFO selects the appropriate readability formulas and presents the results in the results window.

3. Inspecting the results

Formula name	Raw value	Meaning
Document		
Amstad value (AVI): Amstad 1978	39.106	Amstad value: 39 (difficult)
DS cloze forecast (DS): Dickes-Steiner 1977	26.106	DS cloze forecast: 26%
Fucks class (Fucks): Fucks 1955	119.710	Fucks class: 10
Fesse dich kurz index (FDK): Schirm 1971	48.387	FDK index: 48 (diffuse, prolix)
LIX value (G-LIX): LIX für German Bamberger, Vanecek 1984	53.857	LIX class: university
LIX value (G-LIX): LIX für German Bamberger, Vanecek 1984	53.857	LIX class: 9 and higher
RIX value (G-RIX): LIX für German Bamberger, Vanecek 1984	60.000	RIX value: 60
Simplified Measure Of Gobbledygook for German (G-SMOG): McLaughlin 1969	10.048	Simplified Measure Of Gobbledygook Grade: 10
Wheeler-Smith grade (G-WSI): WSI für German Bamberger, Vanecek 1984	48.387	Wheeler-Smith grade: invalid value
PLRI class (PLRI): Beime & Menges 2012	29.218	PLRI class: 29
QU class (QU): Bamberger, Vanecek 1984	10.048	QU class: 10
Tränkle, Bailer index (TB1): readability index Tränkle, Bailer 1984	27.070	Tränkle, Bailer index: 27
TB index (TB2): readability index Tränkle, Bailer 1984	37.840	Tränkle, Bailer index: 37
TRI index (TRI): Text-Redundanz-Index Kuntzsch 1981	N/A	N/A
1. Vienna formula for non-fictional texts (WSTF1): Bamberger, Vanecek 1984	12.451	1. WSTF age: 12
2. Vienna formula for non-fictional texts (WSTF2): Bamberger, Vanecek 1984	12.509	2. WSTF age: 12
3. Vienna formula for non-fictional texts (WSTF3): Bamberger, Vanecek 1984	10.992	3. WSTF age: 10
4. Vienna formula for non-fictional texts (WSTF4): Bamberger, Vanecek 1984	10.922	4. WSTF age: 10
1. Vienna formula for fictional texts (WLTF1): Bamberger, Vanecek 1984	15.828	1. WLTF age: 15
2. Vienna formula for fictional texts (WSTF2): Bamberger, Vanecek 1984	16.333	2. WLTF age: 16
3. Vienna formula for fictional texts (WSTF3): Bamberger, Vanecek 1984	14.311	3. WLTF age: 14
> Paragraph #1		
> Paragraph #2		

Under the text window on the left side are the results of each readability formula and its interpretation. First for the whole document, if you click on a node you will see the results for each paragraph. You can compare the results of the paragraphs to find out which ones are better

understandable and may change the text.

Text metrics: Document

- 31 sentences
- 483 words
- 11 numbers
- 1 other
- 495 all
- 1028 syllables
- 202 words with one syllable
- 336 words with one or two syllables
- 147 words with at least 3 syllables
- 79 words with at least 4 syllables
- 3661 characters
- 3052 letters
- 50 punctuation marks
- 9 words in (Tränkle, Bailer: conjunctions) word list
- 61 words in (Tränkle, Bailer: prepositions) word list
- 7 words in (Text-Redundanz-Index: foreign words) word list
- 235 words not in (WLTF: Bamberger, Vanecek 1984) word list
- 451 words with 3 and more characters
- 227 words with 6 and more characters
- 183 words with 7 and more characters

On the right side you see the text metrics, some values are underlined. If you click on these, all words that fulfil the criterion will be highlighted in yellow colour in the text, e.g. if you want to show all words that have at least six characters, or which words are in the Dale-Chall list of words. You can save the results to the clipboard for further processing – see the file menu.

Transcript of Al Gore's Remarks 50th Anniversary of NATO

We know the inner meaning of what happened here at Ellis Island - where outcasts by the millions, described on the base of this statue as "tired," "poor," "homeless" -- "wretched refuse" -- came in different garb, often isolated by language and culture from the mores they encountered in the land where they arrived. Here, they connected with the American dream.

Here, in the glow of liberty's torch, the forebears of nearly one hundred million Americans first set foot on our shores.

Even on this day of remembrance, Ellis Island is comparatively calm and quiet. But eighty years ago, it was a blur of movement and emotion -- alive with the dreams and fears of more than 5,000 immigrants who arrived on this island every day. Some carrying steamer trunks. Some holding their children close to them. Some coughing from a week's cramped voyage on the open sea.

Many chilled with fear as they faced the health inspectors and customs officials who had the power to welcome them to America, or send them back to where they came from.

Theirs was more than a geographic journey. It was journey of the heart. And in a very real sense, we cherish Ellis Island not merely as a gateway for millions of new Americans, but also as an expression of the most profoundly American values - the values we have pursued in our effort to build that more perfect union: political, economic, and religious freedom; tolerance and diversity; democracy and self-government.

And so, as we honor our immigrants, we also honor the growing dream of people everywhere to create in their own lands a democratic framework through which tolerance and freedom can take root in the human heart.

For fifty years now, NATO has given us the means and the might to defend those values in Europe. And now, after traveling so far on our journey, America and our transatlantic alliance are at a crossroads.

A continent away, hundreds of thousands of homeless Kosovar Albanians are struggling to win back their basic rights and freedoms -- not by leaving their homeland, but by returning to their homeland in peace.

The results of the readability formulas are shown like this:

Readability research has shown that the longer sentences or words are the more difficult they are to understand.

4. Vocabulary module

This module is only available in the professional version. It lists all words, word combinations or word permutations sorted by alphabet or by frequency, ascending or descending. A vocabulary can help you to see errors in orthography, but also in the use of homographs, synonyms, and other interesting aspects of the words used in a text. A vocabulary can also be used as a basis for the construction of a category system in content analysis.

The menu shows a word list with single words, sorted ascending by alphabet. Clicking on the column header changes the sort order (ascending/descending). No word is excluded.

Options for words list generation

Language: English

Number of words: 1

Fixed word order:

Exclude stop words:

Word length limits

minimum: 1

maximum: 60

Hapax legomena:

or specify frequency limits

minimum: 1

maximum: 100000

Generate

The following parameters can be selected:

- language: you can select the language which is important for the sort order of the vocabulary list.
- number of words: 1 means only single words are listed – this is a normal word list. Higher numbers mean the list contains phrases consisting of the number of words specified.
- fixed word order: makes sense only if the number of words is 2 or higher. You can find phrases and their frequency in a text.
- excluding criteria:
 - exclude stop words: words that are in a stop word list are excluded from the list. These lists are dependent on the language and contain articles, pronouns, conjunctions and prepositions.
 - word length limits:
 - * minimum: the minimum length (in characters) a word must have to be in the list
 - * maximum: the maximum length (in characters) a word can have to be in the list
 - or frequency limits
 - * Hapax legomena: words that occur only once are in the list (minimum and maximum frequency are set to 1)
 - * minimum: the minimum frequency a word must have to be listed
 - * maximum: the maximum frequency a word can have to be listed

Lists are generated and sorted either by alphabet – language dependent – or by frequency, ascending or descending. It is possible that some words stored in a list of stop-words are excluded from the output. Stop-words are often words with a high frequency like articles, pronouns, conjunctions, and prepositions. These stop-words lists are language dependent. Other exclusion criteria are the length and/or the frequency of a word.

You can generate simple word lists, list of word sequences where the number of words can be specified by the user, and word permutation which are similar to word sequences, but have no fixed order. The next picture shows you the results of all two-word-frequencies in descending order – the complete list is longer.

Options for words list generation

Language: German

Number of words: 2

Fixed word order:

Exclude stop words:

Word length limits

minimum: 1

maximum: 60

Hapax legomena:

or specify frequency limits

minimum: 1

maximum: 100000

Generate

Freq...	% of all	String
3	0.73	Milliarden Dollar
2	0.49	eine Einigung
2	0.49	den eigenen
2	0.49	Build Back
2	0.49	es in
2	0.49	Das Paket
2	0.49	Back Better
2	0.49	Angaben der
2	0.49	eigenen Reihen
2	0.49	in den
2	0.49	im Senat
2	0.49	dem Namen
2	0.49	der Demokraten
2	0.49	auf ein
2	0.49	Joe Biden
2	0.49	Joe Manchin
2	0.49	Milliarden Euro
2	0.49	mit dem
2	0.49	Präsident Joe

This picture shows a simple word lists, sorted by frequency descending so that the most frequent words occur at the top of the list.

Options for words list generation

Language: German

Number of words: 1

Fixed word order:

Exclude stop words:

Word length limits

minimum: 1

maximum: 60

Hapax legomena:

or specify frequency limits

minimum: 1

maximum: 100000

Generate

Freq...	% of all	String
13	2.62	die
12	2.41	der
11	2.21	das
10	2.01	von
10	2.01	und
8	1.61	zu
8	1.61	auf
7	1.41	den
7	1.41	Demokraten
7	1.41	Manchin
6	1.21	in
6	1.21	US
6	1.21	im
6	1.21	Milliarden
6	1.21	dem
6	1.21	eine
6	1.21	Biden
5	1.01	ein
5	1.01	er

This vocabulary shows all word sequences with a minimum frequency of 2. This list is much shorter than the lists above.

Options for words list generation

Language: German

Number of words: 2

Fixed word order:

Exclude stop words:

Word length limits

minimum: 1

maximum: 60

Hapax legomena:

or specify frequency limits

minimum: 2

maximum: 100000

Generate

Freq...	% of all	String
3	0.73	Dollar Milliarden
2	0.49	anderem unter
2	0.49	Angaben der
2	0.49	auf ein
2	0.49	Back Better
2	0.49	Back Build
2	0.49	Biden Joe
2	0.49	Das Paket
2	0.49	dem mit
2	0.49	dem Namen
2	0.49	Demokraten der
2	0.49	den eigenen
2	0.49	den in
2	0.49	eigenen Reihen
2	0.49	eine Einigung
2	0.49	es in
2	0.49	Euro Milliarden
2	0.49	Flügelkämpfe und
2	0.49	im Senat
2	0.49	Jahre zehn
2	0.49	Joe Manchin
2	0.49	Joe Präsident
2	0.49	Senat und
2	0.49	US von
2	0.49	zehn über