

### TextQuest/ReFo Handbuch

Oktober 2024

 $\mathcal{SSC}$  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

# Inhaltsverzeichnis

1.	Einfi	ührung	5			
	1.1	Basisversion	6			
	1.2	Standardversion	6			
	1.3	Premiumversion	6			
	1.4	Professionelle Version	7			
	1.5	Kombinationen der implementierten Lesbarkeitsformeln	7			
2. Arbeiten mit TEXTQUEST/REFO						
	2.1	Vorbereitung des Textes	9			
	2.2	Sprachauswahl	9			
	2.3	Textgenreauswahl	10			
	2.4	Auswahl der Leserschaft	10			
	2.5	Ergebnisse und ihre Interpretation	11			
3.	Wor	tschatzmodul	13			

# 1. Einführung

Dieses Dokument beschreibt alle Versionen von TextQuest/Refo, wobei ReFo eine Abkürzung für **re**adability **fo**rmulat (Lesbarkeitsformel) ist. Diese Software ist für MS-Windows und für Apple MacOS erhältlich. Alle Versionen sind für verschiedene Lizenzzeiten und verschiedene Benutzer verfügbar. Ein Text wird analysiert, die Ergebnisse einer Analyse sind Werte verschiedener Lesbarkeitsformeln.

In den letzten 100 Jahren wurde eine Vielzahl von Lesbarkeitsformeln entwickelt. Sie alle basieren auf einem Lückentextverfahren, das Ergebnis dieser Analyse ist eine mathematische Formel, die auf einer Regressionsanalyse beruht. In den meisten Formeln sind die Länge der Sätze und die Länge der Wörter die wichtigsten Variablen, die die Lesbarkeit eines Textes erklären. Lesbarkeitsformeln wurden für verschiedene Sprachen, verschiedene Textgattungen und verschiedene Leser entwickelt - daher müssen Sie die richtige Formel kennen und auswählen. TextQuest/ReFo kennt jede Kombination und wählt die passende(n) Formel(n) für Ihre Bedürfnisse aus.

Die Ergebnisse hängen von der/den Lesbarkeitsformel(n) ab und werden für den gesamten Text und für alle seine Absätze berechnet. So können Sie sehen, welche Absätze besser oder schlechter sind als andere. Die Ergebnisse können sein:

- ein Indexwert zwischen 0 und 100 je höher der Wert, desto besser ist die Lesbarkeit des Textes. Es wird auch eine verbale Beschreibung gegeben, damit Sie den Wert interpretieren können.
- Lesealter: das Alter des Empfängers, der in der Lage sein muss, den Text zu verstehen. Dies hängt mit dem Schulsystem in den USA zusammen.
- Leseklasse: Die Schulklasse des Lesers muss in der Lage sein, den Text zu verstehen. Dies steht im Zusammenhang mit dem Schulsystem der USA.

TextQuest/ReFo unterstützt:

- Sprachen: Englisch, Französisch, Deutsch, Spanisch, Italienisch, Niederländisch, Flämisch, Afrikaans, Türkisch, Polnisch, Dänisch, Norwegisch, Isländisch, Schwedisch, Tschechisch, Slowakisch, Slowenisch.
- Textgattung: unspezifisch, fiktional, nicht-fiktional, technische Handbücher, Kinderbücher, Schulbücher, politische Texte, Packungsbeilagen von Medikamenten (Beipackzettel)

• Leserschaft: unspezifisch, Kinder, Jugendliche, Erwachsene, ausländische Lernende, Soldaten, Patienten.

Nicht alle Sprachen, Textgattungen und Leser werden von jeder Version unterstützt. So ist es möglich, dass Sie nur für eine Lizenz zahlen, die das beinhaltet, was Sie benötigen.

### 1.1 Basisversion

Die Basisversion deckt nicht alle Sprachen, alle Textgattungen oder alle Leserschaften ab. Die folgenden Funktionen werden unterstützt:

- Sprachen: Englisch, Deutsch
- Textgattung: unspezifisch, technisch, fiktional, nicht-fiktional, Kinderbücher, Schulbücher, politische Texte
- Leserschaft: unspezifisch, Kinder bis 10 Jahre, Jugendliche von 11 bis 18 Jahren, Soldaten

#### 1.2 Standardversion

Diese Version deckt die 14 Sprachen ab, enthält aber nicht das Wortschatzmodul wie die Professionalversion.

- Sprachen: Englisch, Deutsch, Französisch, Spanisch, Italienisch, Niederländisch, Flämisch, Afrikaans, Dänisch, Dänisch, Isländisch, Norwegisch, Schwedisch, Türkisch
- Textgattungen: unspezifisch, technisch, fiktional, nicht-fiktional, Kinderbücher, Schulbücher, politische Texte, Packungsbeilagen/Beipackzettel von Medikamenten
- Leserschaft: unspezifisch, Kinder bis zu 10 Jahren, Jugendliche von 11 bis 18 Jahren, Soldaten, Fremdsprachenlerner

### 1.3 Premiumversion

Diese Version deckt alle Sprachen ab, enthält aber nicht das Wortschatzmodul wie die Professionalversion.

• Sprachen: Englisch, Deutsch, Französisch, Spanisch, Italienisch, Niederländisch, Flämisch, Afrikaans, Dänisch, Finnisch, Dänisch, Isländisch, Norwegisch, Polnisch, Portugiesisch, Tschechisch, Slowakisch, Kroatisch, Slowenisch, Schwedisch, Türkisch

- Textgattungen: unspezifisch, technisch, fiktional, nicht-fiktional, Kinderbücher, Schulbücher, politische Texte, Packungsbeilagen/Beipackzettel von Medikamenten
- Leserschaft: unspezifisch, Kinder bis zu 10 Jahren, Jugendliche von 11 bis 18 Jahren, Soldaten, Fremdsprachenlerner

### 1.4 Professionelle Version

Professional: wie die Standardversion, aber mit einem zusätzlichen Vokabelmodul. Dieses zeigt Wortlisten, alle Wortkombinationen je nach Anzahl der Wörter oder alle Permutationen innerhalb eines Satzes (welche Wörter kommen zusammen vor).

### 1.5 Kombinationen der implementierten Lesbarkeitsformeln

TEXTQUEST/REFO wählt die Lesbarkeitsformeln aus fr viele Kombinationen von Sprach, Textgenre und Leserschaft. Für bestimmte Kombinationen gibt es allerdings keine Lesbarkeitformeln. Aber für jede Sprache gibt mindestens eine universelle Lesbarkeitsformel für alle Texte und/oder alle Leserschaften. Die implmentierten Kombinationen zeigt die folgende Tabelle:

Sprache	Kombination			
portugiesisch	Kinder älter als 10 Jahre			
holländisch	nicht fiktionale Texte und Kinderbücher für Kinder unter 10 Jahren (2)			
	alle Texte und fiktionale Texte für Kinder älter als 10 Jahre (2)			
polnisch	nicht fiktionale Texte			
türkisch	Schulbücher für Kinder unter 10 Jahren (2)			
englisch	Prosa für Soldaten (2)			
	technische Texte für Soldaten (5)			
	Prosa für Kinder älter als 10 Jahre (2)			
	Schulbücher für Kinder älter als 10 Jahre (3)			
	Kinderbücher für Kinder älter oder jünger als 10 Jahre(3)			
	Schulbücher für Kinder unter 10 Jahren (2)			
	Schulbücher für alle Altersgruppen			
deutsch	politische Texte für Erwachsene			
	alle Texte für Kinder unter 10 Jahren			
	fiktionale Texte für Kinder älter als 10 Jahre (2)			
	nicht-fiktionale und fiktionale Texte für Kinder unter 10 Jahren (8)			

Lesbarkeitsformeln für Beipackzettel gibt es für folgende Sprachen: Dänisch, Deutsch, Englisch, Finnisch, Französisch, Isländisch, Norwegisch, Slowenisch, Spanisch und Schwedisch.

# 2. Arbeiten mit TEXTQUEST/REFO

Nach dem Start sehen Sie das große Editorfenster. Sie können Ihren Text in dieses Fenster einfügen. Wenn Ihr Text in einer Datei steht, benutzen Sie das Dateimenü, um den Text zu importieren.

### 2.1 Vorbereitung des Textes

In den meisten Fällen genügt es, den Text auszuschneiden und in das Bearbeitungsfenster einzufügen. Wenn Sie HTML-, PHP- oder Java(Skript)-Code im Text haben, müssen Sie diesen löschen, bevor Sie eine Analyse starten.

Eine Lesbarkeitsanalyse besteht aus dem Zählen von grammatikalischen Sätzen und Silben. TEXT-QUEST/REFO zerlegt den gesamten Text in grammatikalische Sätze und zählt die Silben. Sie können den Text in Absätze aufteilen – trennen Sie die Absätze einfach durch eine Leerzeile. Wenn Sie eine Überschrift (ohne .?! am Ende) haben, machen Sie einen Absatz daraus, indem Sie eine Leerzeile einfügen.

Viele Abkürzungen, die Punkte verwenden, sind TEXTQUEST/REFO bekannt. Wenn nach einer Abkürzung ein neuer Satz begonnen wird, entfernen Sie einfach den Punkt. Es ist sehr unwahrscheinlich, dass Sie dies jemals tun müssen, da viele Abkürzungen – je nach Sprache – in internen Listen gespeichert sind und nicht als Satzende, sondern als Abkürzung erkannt werden.

Vor allem in Beipackzetteln gibt es Aufzählungen, die sehr lange Sätze ergeben und die Ergebnisse verfälschen können. Daher können Sie jeden Aufzählungsteil mit einem - (Minuszeichen) am Anfang jeder Zeile beginnen, die ein Aufzählungselement enthält.

### 2.2 Sprachauswahl

In der Mitte des Bildschirms können Sie die Sprache auswählen. Bitte beachten Sie, dass Sie in der Basisversion nicht alle Sprachen auswählen können. Abkürzungen sind sprachspezifisch, einige tausend sind in TEXTQUEST/REFO implementiert. Im Vokabularmodul wird in der Profiversion auch die korrekte Sortierreihenfolge der gewählten Sprache verwendet, z.B. wird im Deutschen der Umlaut ä wie sein Grundvokal a behandelt, während im Dänischen ä hinter dem z einsortiert wird.

anguage:	English	Genre: All	Reader: All	~
	Czech	art Analysis	Clear Besults	
	Danish	IT Analysis	Clear Results	
	German			
	English			
	Spanish			
	Finnish			
	French			
	Croatian			
	Icelandic			
	Italian			
	Dutch, Flamish, Afrikaans			
	Norwegian			
	Polish			
	Portuguese			
	Slovak			
	Slovens			
	Sidverie			
	Sweaish			
	Turkish			

### 2.3 Textgenreauswahl

Language:	English	~	Genre:	All	~	Reader:	All	~
		Sta	art Ana	All Technical texts		s		
				Fictional texts Non fictional texts Childrens books School books Political texts Patient leaflet inform	ation			

Die zweite Option ist die Auswahl des Textgenres, *all* bedeutet unspezifisch – gilt für alle Texte. Beipackzettel können mit der Basisversion nicht analysiert werden. Viele Lesbarkeitsformeln wurden für spezielle Textgattungen entwickelt, die Ergebnisse sind nicht auf andere Textgattungen übertragbar. Manchmal gilt auch eine Kombination aus Textgattung und Leserschaft, z.B. bei fiktionalen Texten und Kinderbüchern.

### 2.4 Auswahl der Leserschaft

Die letzte Option besteht darin, auszuwählen, ob es sich um eine bestimmte Leserschaft handelt, oder, falls dies nicht der Fall ist, *all* zu wählen. Einige Lesbarkeitsformeln sind eine Kombination aus Textgattung und Leserschaft, TEXTQUEST/REFO wählt die entsprechenden Formeln aus. Es sind keine weiteren Optionen verfügbar

#### 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

Sie können einige unterstrichene Objekte anklicken und so alle Wörter mit mehr als 3 Silben, Wörter, die in einer Liste enthalten sind oder nicht enthalten sind, sehen. Nach der Analyse können Sie lange Sätze oder Wörter markieren und festlegen, wie viele Wörter oder Buchstaben als zu lang angesehen werden. Diese werden in gelb angezeigt.



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

Formula name	Raw value	Meaning
✓ Document		
LIX value (G-LIX): LIX für German Bamberger, Vanecek 1984	53.469	LIX class: university
✓ Paragraph #1		
LIX value (G-LIX): LIX für German Bamberger, Vanecek 1984	65.545	LIX class: university
✓ Paragraph #2		
LIX value (G-LIX): LIX für German Bamberger, Vanecek 1984	39.366	LIX class: university
✓ Paragraph #3		
LIX value (G-LIX): LIX für German Bamberger, Vanecek 1984	53.705	LIX class: university
✓ Paragraph #4		
LIX value (G-LIX): LIX für German Bamberger, Vanecek 1984	68.652	LIX class: university
✓ Paragraph #5		
LIX value (G-LIX): LIX für German Bamberger, Vanecek 1984	45.833	LIX class: university
✓ Paragraph #6		
LIX value (G-LIX): LIX für German Bamberger, Vanecek 1984	68.637	LIX class: university
✓ Paragraph #7		
LIX value (G-LIX): LIX für German Bamberger, Vanecek 1984	49.676	LIX class: university
✓ Paragraph #8		
LIX value (G-LIX): LIX für German Bamberger, Vanecek 1984	51.000	LIX class: university
✓ Paragraph #9		
LIX value (G-LIX): LIX für German Bamberger, Vanecek 1984	59.238	LIX class: university
✓ Paragraph #10		
LIX value (G-LIX): LIX für German Bamberger, Vanecek 1984	47.333	LIX class: university

Die Ergebnisse der Lesbarkeitsformeln werden wie folgt angezeigt:

Die Lesbarkeitsforschung hat gezeigt, dass Sätze oder Wörter umso schwerer zu verstehen sind, je länger sie sind.

Wenn Sie auf einen Knoten klicken, werden Ihnen die Ergebnisse für diesen Absatz angezeigt. So können Sie die Ergebnisse der einzelnen Absätze vergleichen, um Unterschiede zu erkennen und eine Vorstellung davon zu bekommen, welcher Teil des Textes verbessert werden kann.

## 3. Wortschatzmodul

Dieses Modul ist nur in der Professionalversion verfügbar. Es listet alle Wörter, Wortkombinationen oder Wortpermutationen auf, sortiert nach Alphabet oder nach Häufigkeit, auf- oder absteigend. Ein Vokabular kann Ihnen helfen, Fehler in der Rechtschreibung, aber auch die Verwendung von Homographen, Synonymen und andere interessante Aspekte der in einem Text verwendeten Wörter zu erkennen. Ein Vokabular kann auch als Grundlage für den Aufbau eines Kategoriensystems in der Inhaltsanalyse verwendet werden.

Das Menü zeigt eine Wortliste mit einzelnen Wörtern, aufsteigend nach dem Alphabet sortiert. Ein Klick auf die Spaltenüberschrift ändert die Sortierreihenfolge (aufsteigend/absteigend). Kein Wort wird ausgeschlossen.

Options for words list gener	ation	
Language:	English	~
<u>N</u> umber of words:	1	-
<u>Fixed word order:</u>		
Exclude stop words:		
Word length limits		
minimum:	1	▲ ▼
maximum:	60	▲ ▼
Hapax legomena:		
or specify frequency limits		
minimum:	1	
maximum:	100000	
Ger	nerate	

Die folgenden Parameter können ausgewählt werden:

- Sprache: Sie können die Sprache auswählen, die für die Sortierreihenfolge der Vokabelliste wichtig ist.
- Anzahl der Wörter: 1 bedeutet, dass nur einzelne Wörter aufgelistet werden -dies ist eine normale Wortliste. Höhere Zahlen bedeuten, dass die Liste Phrasen enthält, die aus der angegebenen Anzahl von Wörtern bestehen.
- feste Wortfolge: macht nur Sinn, wenn die Anzahl der Wörter 2 oder höher ist. Sie können Phrasen und deren Häufigkeit in einem Text finden.
- Ausschlusskriterien:
  - Ausschließen von Stopwörtern: Wörter, die in einer Stopwortliste stehen, werden von der Liste ausgeschlossen. Diese Listen sind von der Sprache abhängig und enthalten beispielsweise Artikel, Pronomen, Konjunktionen und Präpositionen.
  - Wortlängengrenzen:
    - \* minimum: die Mindestlänge (in Zeichen), die ein Wort haben muss, um in der Liste zu stehen
    - \* maximum: die maximale Länge (in Zeichen), die ein Wort haben darf, um in der Liste enthalten zu sein
  - Häufigkeitsgrenzen
    - \* Hapax legomena: Wörter, die nur einmal vorkommen, sind in der Liste (minimale und maximale Häufigkeit sind auf 1 gesetzt)
    - \* minimum: die Mindesthäufigkeit, die ein Wort haben muss, um in der Liste zu stehen
    - \* maximum: die maximale Häufigkeit, die ein Wort haben darf, um in der Liste zu sein

Die Listen werden entweder nach dem Alphabet – abhängig von der Sprache – oder nach der Häufigkeit, auf- oder absteigend, sortiert. Es ist möglich, dass einige Wörter, die in einer Liste von Stop-Wörtern gespeichert sind, von der Ausgabe ausgeschlossen werden. Bei den Stopwörtern handelt es sich oft um Wörter mit einer hohen Häufigkeit wie Artikel, Pronomen, Konjunktionen und Präpositionen. Diese Stopwortlisten sind sprachabhängig. Weitere Ausschlusskriterien sind die Länge und/oder die Häufigkeit eines Wortes. Sie können einfache Wortlisten, Listen von Wortfolgen, bei denen die Anzahl der Wörter vom Benutzer festgelegt werden kann, und Wortpermutationen erzeugen, die Wortfolgen ähneln, aber keine feste Reihenfolge haben. Das nächste Bild zeigt Ihnen die Ergebnisse aller Zwei-Wort-Häufigkeiten in absteigender Reihenfolge — die vollständige Liste ist länger.

Options for	r words list	generation		
	Lang	uage: German	~	
Nu	umber of w	vords: 2	-	
Fi	xed word o	order: 🔽	- Canad	
Excl	ude stop w	vords:		
We	ord length	limits		
	minin	mum: 1		
	maxin	mum: 60	•	
Ha or specify f	apax legon frequency l minir maxir	nena:  iimits mum:  1 *  num:  100000 *		
	(	Generate		
Freq v	% of all	String	T	
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	49 der Demokraten		
2	0.49	9 auf ein		
2	0.49	Joe Biden		
2	0.49	Joe Biden Joe Manchin		
2 2 2	0.49 0.49 0.49	Joe Biden Joe Manchin Milliarden Euro		
2 2 2 2	0.49 0.49 0.49 0.49	Joe Biden Joe Manchin Milliarden Euro mit dem		

Dieses Bild zeigt eine einfache Wortliste, die nach Häufigkeit absteigend sortiert ist, so dass die häufigsten Wörter am Anfang der Liste stehen.

Options for	words list	generation	
	Lang	age: German	~
Nu	mber of w	ords: 1	
Fi	ed word o	rder:	(heread)
Exclu	ude stop w	ords:	
We	rd length i	imits	
	minin	1	
	maxin	num: 60	-
Ha or specify f	apax legon requency l	nena: 🔲 imits	
	minin	num: 1 📮	
	mavin	100000	
	97 af all	Generate	
Freq V	76 OT 81	string	
13	2.02	die	
11	2.41	der	
10	2.01	VOD	
10	2.01	und	
8	1.61	20	
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	

Dieses Vokabular zeigt alle Wortfolgen mit einer Mindesthäufigkeit von 2. Diese Liste ist viel kürzer als die obigen Listen.

Options for	words list	gene	ation		
	Lang	uage:	German	~	
Nu	mber of w	ords:	2	1	
Fig	xed word o	orden			
Exclu	ude stop w	ords:			
11/-	and Laura while 1	in the			
wo	autengun	irrines	1		
	minir	num:			
	maxin	num:	60	÷	
Ha or specify f	apax legon requency	nena: limits	0		
	minir	num:	2		
	maxir	num:	100000 📮		
	6				
	l	Gei	herate		
Freq 🔻	% of all	Strin	g		
3	0.73	Dolla	ar Milliarden		
2	0.49	ande	erem unter		
2	0.49	Ang	aben der		
2	0.49	auf e	sin		
2	0.49	Back	Better		
2	0.49	Back	: Build		
2	0.49	Bide	n Joe		
2	0.49	Das	Paket		
2	0.49	dem	mit		
2	0.49	dem Namen			
2	0.49	Dem	okraten der		
2	0.49	den	eigenen		
2	0.49	den	in		
2	0.49	eige	nen Reihen		
2	0.49	eine	Einigung		
2	0.49	es in	6		
2	0.49	Euro Milliarden			
2	0.49	Flüg	Flügelkämpfe und		
2 0.49 im		im S	im Senat		
2	0.49	Jahr	e zehn		
2	0.49	Joel	Manchin		
2	0.49	Joef	Präsident		
2	0.49	Sena	t und		
2	0.49	USV	on		
2	0.49	zehn	über		

Das alte TextQuest hat dieses Problem durch die Verwendung verschiedener Dateierweiterungen in seinem automatischen Dateibenennungssystem gelöst. Ausschluss von Wörtern: entweder vor dieser Analyse, oder der Benutzer kann diese Optionen auch in dieser Aufgabe angeben.