Addressing Readability in NL Requirements Analysis

Gianluca Trentanni

(gianluca.trentanni@isti.cnr.it)

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", CNR, Pisa, Italy

Abstract. What is "*readability*" and why does it matter? According to Wikipedia (2006), "*Readability is a measure of the accessibility of a piece of writing, indicating how wide an audience it will reach. Readability is a judgment of how easy a text is to understand*". In the following, the use of Readability Metrics in automated analysis of Natural Language expressed Software Requirements is investigated and some use cases are presented.

Keywords: Software Engineering, Requirements Analysis, Metrics, Requirements/Specifications, Metrics

Introduction

What is "readability" and why does it matter? According to Wikipedia (2006), "Readability is a measure of the accessibility of a piece of writing, indicating how wide an audience it will reach. Readability is a judgment of how easy a text is to understand".

Requirements and specifications documents have become the must have of the modern software project. Companies and software houses finally see them as both instruments that help to capture business rules and direct links to financial success or liability.

A good requirements writing not only educate but also keep all parties of the software development process well and clearly informed throughout the project life cycle. Poor or vague requirements and inconsistent elicitation processes often is the cause of the development of undesirable products with bugs or functionality that does not meet the business need and might waste it.. A flawed system production could lead to significant financial exposure and possibly even the loss of significant sums of money to a provider or to the client or to both.

Starting from version 0.4, QuARS Express [18, 19, 20, 21] (a tool for the automatic quality analysis of Natural Language (NL) software requirements [13]) shows the readability analysis of sentences/requirements. This new feature exploits the GNU program called "Diction/Style" [1] referenced on page

http://www.gnu.org/software/diction/diction.html.

The seven indices which are calculated, Kincaid [2, 3, 4], ARI [5], Coleman-Liau [6], Flesch [7, 8, 9, 10], FOG [11], LIX [12, 13, 14, 15], SMOG [16]) are well known in the readability research field.

The Style program analyzes the surface characteristics of the writing style of a document and prints various readability grades and length of words, sentences, and paragraphs.

This article tries to resume and explain how to interpret the several indices and formulas used, identifying their average application ranges and their average values.

Most of the seven indices (indicated with a GL in the table below) aren't real indexes, but rather grade level specifications. In this case it is clearly hard to define a real "*better*" value because the lower is the value the simpler is the sentence and the more readable is the requirement. For these level grades we will refer to a medium range of instruction grades, assuming as "*ideal*" value its average. Anyway the average values which are pointed out are higher than common every-day text scores and it's advisable to keep the writings under these values.

In the explanation chapter, some examples are provided to ease the comparison and several every-day common readability scores are listed as well.

Finally some applications of QuARS to requirements from the real world of industry are shown.

Readability Indexes Ranges Summary

The following table (**Table 1**) presents, just as an indication, the best application ranges (with averages) of the used readability indexes.

Index	Kind	Range	Average
Kincaid	**GL**	**5,5 --> 16,3** (The **lower** the score the more readable the text)	**11**
ARI	**GL**	**9 --> 16** (The **lower** the score the more readable the text)	**12,5**
Coleman-Liau	**GL**	**10 --> 15** (The **lower** the score the more readable the text)	**12,5**
Flesch	**I**	**60 --> 70** (The **higher** the score the more readable the text)	**65**
FOG	**GL**	**10 --> 15** (The **lower** the score the more readable the text)	**12,5**
LIX	**I**	**34 --> 58** (the **lower** the score the more readable the text)	**46**
SMOG	**GL**	**10 --> 15** (The **lower** the score the more readable the text)	**12,5**
(Where "GL" stands for "Grade Level" and refers to a US school grade and "I" stands simply for "Index")			

Table 1: best application ranges (with averages) of some readability indexes

The average indicates the recommended value of the index for a wider readability and understandability of the examined text.

Higher and lower values are possible anyway, and information on the better or worse readability that they point out rely on the related index formula and its interpretation.

So, what's a good score? A "*good*" score, of course, depends on your target audience, your chosen style, and what you mean to say.

While the Grade Level seems arbitrary depending on what your target audience is, it should be said that the average best-seller is around Grade 4 level, newspapers about 6, and business books 7-8. Lower than you might have thought. Probably it is impossible reach such low values for a technical document, anyway if you want to make your sentence (or requirement) easy to read, well understandable and unambiguous, the better value is likely the lowest grade level possible.

Some more details are shown in the chapter "Readability Indexes Explained".

Moreover, some samples of average values in every-day documents are shown in the chapter "Common Readability scores".

Readability Indexes Formulas Summary

Every readability grade or index relies on a formula related to word count, syllabes count, complex words count and so on.

There are dozens of readability formulas, more or less sophisticated, all based on the same concepts.

Index	Formula
Kincaid	**GL = 11.8*syllables/words+0.39*words/sentences-15.59**
ARI	**GL = 4.71*chars/words+0.5*words/sentences-21.43**
Coleman-Liau	**GL = 5.89*chars/words-0.3*sentences/(100*words)-15.8**
Flesch	**I = 206.835-84.6*syllables/words-1.015*words/sent**
FOG	**GL = 0.4*(words/sent+100*((words >= 3 syllables)/words))**
LIX	**I = words/sent+100*(words >= 6 char)/words**
SMOG	**GL = square root of (((words >= 3 syllables)/sent)*30) + 3**
(Where "GL" stands for "Grade Level" and refers to a US school grade and "I" stands simply for "Index")	

Table 2: Formulas specifications for some readability indexes

If you use short, average sentence lengths and few big words, you can reduce the reading level and increase the speed and ease of reading your writing. Although based on proven research, readability results should be considered merely as a guideline and reference point to ease the readability of writings.

Table 2 shows the formulas used by the Style tool, which are the more widely known and used readability aids.

Readability Indexes Explanation

What does a "*Reading Grade Level"* or "*Index"* mean?

The reading grade level of a text depends on the use of the text. If the text is used for independent, unassisted, or recreational use, the reading grade level will be higher than the one related to a text destined for classroom use and optimum learning gain. In other words, the same text will be easier for those with more advanced reading skills (with a higher grade level) and harder for those with less (and with a lower grade level).

In the next paragraphs we will see more in detail the explanation of the seven readability indexes that QuARS Express uses: Kincaid, ARI, Coleman-Liau, Flesch, FOG, LIX, , MOG.

Flesch-kincaid Readability Grade Level

The Flesch-Kincaid (also known as "Flesch Kincaid Grade Level" and simply referred as "Kincaid" in this report) has been around for 50 years and analyzes a sample of writing by examining the number of words, syllables and sentences. This readability grade level can be found even in Microsoft Word and Word Perfect and can analyze writings in seconds. The Flesch-Kincaid formula has been developed for Navy training manuals that ranged in difficulty from 5.5 to 16.3. It is probably best applied to technical documents because it is based on adult training manuals rather than school book text. On the other hand, scientific texts with many long scientific terms are rated higher, although they are not necessarily harder to read for people who are familiar with those terms. It's a refinement to the Flesch Index and its score analyzes and rates text on a U.S. grade-school

level based on the average number of syllables per word and words per sentence. For example, a score of 8.0 means that an eighth grader would understand the text and text with a Flesch-Kincaid score of 10.1 would be considered suitable for someone with a 10th grade or higher reading level. Given standard writing averages seventh to eighth grade, aim for a Flesch-Kincaid score between 7.0 and 8.0. This test, along with Simplified ARI and New Fog Count, is part of the Navy Readability Indices collection of tests.

Its average can be pointed to 11, but, clearly, the lower the score the more readable the text.

Automated Readability Index (ARI)

The Automated Readability Index (shortly ARI or "auto") is designed to gauge the understandability of a text. Like many of the other indexes, its output is an approximate representation of the U.S. grade level needed to comprehend the text. It was originally created for U.S. Air Force materials and was designed for technical documents and manuals. Unlike the other indices, ARI, along with Coleman-Liau, relies on a factor of characters per word, instead of the usual syllables per word. ARI is typically higher than Kincaid and Coleman-Liau, but lower than Flesch.

Its average can be pointed to 12.5 but, clearly, the lower the score the more readable the text.

Coleman-Liau Readability Grade Level

The Coleman-Liau readability formula calculates the U.S. grade level of a text sample based on sentence lengths and number of characters. This test usually yields the lowest grade when applied to technical documents. The "Coleman-Liau" Formula usually gives a lower grade than the ones yielded by Kincaid, Automated Readability Index, and Flesch (when applied to technical documents).

Its average can be pointed to 12.5 but, clearly, the lower the score the more readable the text.

Flesch Reading Ease Readability index

The Flesch reading easy formula was developed by Flesch in 1948, and it is based on school texts covering grade 3 to 12. It is wide spread, especially in the USA, because of good results and simple computation. The index is usually between 0 (hard) and 100 (easy); standard English documents average approximately 60 to 70

The Flesch Reading Ease readability score formula rates text on a 100-point scale based on the average number of syllables per word and words per sentence. The higher the Flesch Reading Ease score, the easier it is to understand the document. The obtained Flesch reading Ease readability score is mapped to the corresponding readability level based on the value scale presented in Table 3, where it is shown that its average can be pointed to 65. In this case, anyway the higher the score the more readable the text.

Flesh Score mapping table

Flesch Reading Ease Score		Readability Level
0 - 29	->	Very difficult
30 - 49	->	Difficult
50 - 59	->	Fairly difficult
60 - 69	->	Standard
70 - 79	->	Fairly easy
80 - 89	->	Easy
90 - 100	->	Very easy

Table 3: Flesch Score mapping table

Gunning's FOG Readability Index

The Gunning fog index, developed by Robert Gunning in 1952, is a test designed to measure the readability of a sample of English writing. The resulting number is an indication of the number of years of formal education that a person requires in order to easily understand the text on the first reading. That is, if a passage has a FOG index of 12, it has the reading level of a U.S. high school senior. It is one of the simplest and most effective manual tools for analyzing readability. Gunning defines hard words as those with more than two syllables. To get to a fourth-grade readability level, you need to write with an average sentence length of eight words and no more than one out of 50 words being three or more syllables. It is relatively easy to calculate and accurate within one grade level. The ideal score for readability with the Fog index could be 7 or 8; anything above 12 is too hard for most people to read. For reference, The Bible,

Shakespeare and Mark Twain all have Fog Indexes of about 6. The New York Times has an average of 11-12, Time Magazine about 11. Typically, technical documentation has a FOG Index between 10 and 15, and professional prose almost never exceeds 18.

Average value for technical documents can be pointed to 12.5 but, clearly, the lower the score the more readable the text.

Laesbarhedsindex (LIX) Readability Index

The LIX formula developed by Björnsson from Sweden is very simple and employs a mapping table related to the school year:

LIX Index	...	34		38		41		44		48		51		54		57	...
School year	...		5		6		7		8		9		10		11	...	

Table 4: LIX School Years mapping table

The LIX formula checks for long words (with more then 7 characters) and outputs a number from "*very easy*" (0-34: below school year 5) to "*difficult*" (over 54: above school year 9).

Its average can be pointed to about **46** but, clearly, the **lower** the score the more readable the text.

SMOG-Grading Readability Grade Level

The SMOG Readability Formula is a simple method you can use to determine the reading level of your written materials. Unlike any of the

other formulas, SMOG predicts the grade level required for 100% comprehension. This means that if a person reads at or above a grade level, they will understand 90-100% of the information.

SMOG Grade	Education Level	Example
0 - 6	low-literate	Soap Opera Weekly
7	junior high school	True Confessions
8	junior high school	Ladies Home Journal
9	some high school	Reader's Digest
10	some high school	Newsweek
11	some high school	Sports Illustrated
12	high school graduate	Time Magazine
13 - 15	some college	New York Times
16	Education Level	Atlantic Montly
17 - 18	post-graduate studies	Harvard Businness Review
19+	post-graduate degree	IRS Code

Table 4: SMOG Readability Indexes and ranges

Generally, you should aim for a reading level of sixth grade or less. SMOG Grades 13-16 indicate the need for college education, 17-18 the need for graduate training, while 19 and above the need for a higher

professional qualification. The following table maps the education level to the SMOG calculation results with some example.

Its average can be pointed to 12.5 but, clearly, the lower the score the more readable the text.

Common Readability Scores

In the following table, some common web sites scores are shown in order to enable a quick comparison and point out how they can differ depending on the wording.

NY Times, article

- Kincaid: 6.2
- ARI: 6.2
- Coleman-Liau: 11.8
- Flesch Index: 70.4
- Fog Index: 8.9
- Lix: 34.1 = school year 5
- SMOG-Grading: 9.1

PC World, article

- Kincaid: 10.6
- ARI: 12.1
- Coleman-Liau: 11.9
- Flesch Index: 59.4

Whitehouse Press Release

- Kincaid: 4.1
- ARI: 3.7

- Coleman-Liau: 9.1
- Flesch Index: 84.4
- Fog Index: 7.6
- Lix: 28.3 = below school year 5
- SMOG-Grading: 8.2
- Fog Index: 13.3
- Lix: 47.1 = school year 8
- SMOG-Grading: 11.4

Intuitive.com home page

- Kincaid: 9.7
- ARI: 10.9
- Coleman-Liau: 11.0
- Flesch Index: 64.0
- Fog Index: 12.8
- Lix: 45.1 = school year 8
- SMOG-Grading: 11.2

ESPN.Go.com article

- Kincaid: 12.3
- ARI: 6.7
- Coleman-Liau: 15.6
- Flesch Index: 17.4
- Fog Index: 15.7
- Lix: 39.3 = school year 6
- SMOG-Grading: 10.7

Nickelodean Home Page

- Kincaid: 4.0
- ARI: 3.5

- Coleman-Liau: 9.1
- Flesch Index: 84.2
- Fog Index: 6.2
- Lix: 25.0 = below school year 5
- SMOG-Grading: 7.0

QuARS *Express* and readability analysis

The tool QuARS (Quality Analyzer for Requirements Specifications), analyzes automatically requirement documents pointing out linguistic ambiguities and easing the cleanse activity of the Requirement Engineer. In a few words, it performs a lexical analysis of requirements detecting ambiguous terms or wordings.

A modified version of QuARS, namely QuARS *Express,* has been developed to handle a more complex and structured data format containing metadata (requirements authors, belonging project or subproject and so on.) and to produce an analysis report richer of categorized information. The information grows as a function of the number of metadata items available and the size of the report grows consequently and can be composed of several pages. As an improvement of the simple text based report made by QuARS the new report exploits the HTML technology to produce structured hypertextual pages.

As a set of new metrics, QuARS *Express* introduces readability analysis, by means of some of the existing readability indexes, both concerning the whole requirements document and the single requirement it contains, exploiting the well know tool developed by the GNU crew, called "*Diction/ Style"*.

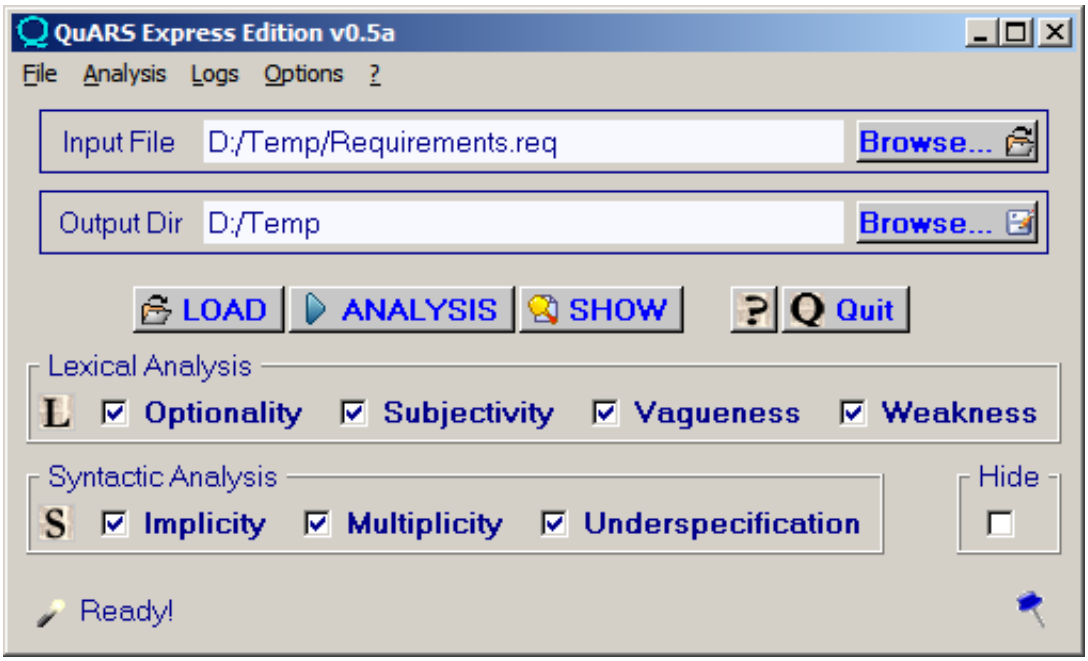

Fig. 1 - QuARS *Express* GUI

Readability Analysis and Requirements: a use case

The tool QuARS *Express* (Fig. 1) has been used in several european projects and in other collaborative activities with third industrial parties.

MODCONTROL is the european project during which the basic idea of introducing the readability analysis has been developed. EU/IP MODTRAIN project, subproject MODCONTROL [22], addresses the standardization of an innovative Train Control and Monitoring System (TCMS) system, due to the capability to point out potential sources of ambiguous definitions and other weaknesses.

During MODCONTROL's specification phase, project partners have gathered requirements from different documents sources such as specifications of existing trains, standards or drafted specifications from other EU projects. These requirements have then been consolidated,

harmonized and refined among the project partners in several review sessions. The final requirements document generated is composed of about 5700 requirements expressed as Natural Language and categorized as Functional Requirements (FREQ) , System Requirements (SREQ)

Readability Index	FREQ Scores	SREQ Scores
Kincaid	13.5	7.4
ARI	15.6	7.6
Coleman Liau	14.2	13
Flesch Index	44.8/100	63.4/100
Fog Index	16.8	10.4
LIX	56.5	40.7
SMOG-Grading	14.2	10.1

Table 5 - MODCONTROL Project, Readability analysis results

Along with the other analysis more properly linguistic, sintactical and lexical, the readability analysis has been performed on the document as well.

Table 5 shows the readability average scores of the two documents, FREQ and SREQ. Note that the SREQ document results to be more readable than the FREQ one. In fact, the indexes values of the SREQ document stand in reasonable ranges according to their technical nature, whereas the scores of the FREQ document are higher than we expected.

Indeed, values of the Kincaid, ARI, Coleman-Liau, FOG, SMOG indexes higher than 15, of the LIX index higher than 58, and of the Flesh index lower than 60 give the indication of a hardly readable document. In our case FREQ exceeds most of such indexes, and it is close to the limits

for the other ones: though this is not a dramatic defect, it is advisable to improve the readability of functional requirements, for example shortening phrases and splitting paragraphs.

References

[1] http://www.gnu.org/software/diction/diction.html

[2] Kincaid, J.P., Fishburne, R.P, Rogers, R.L. & Chissom, B.S. "*Derivation of new readability formulas (automated readability index, Fog count and Flesch reading ease formula) for navy enlisted personnel*" in Research Branch Report 8-75. Naval Air Station, Memphis, TN, 1975

[3] Farr, J. N., Jenkins, J. J., and Paterson, D. G. "*Simplification of Flesch Reading Ease Formula*" in Journal of Applied Psychology, Volume 35, Number 5, (October), pp. 333-337, 1951

[4] Kincaid, J. P. & McDaniel, W.C. "*An inexpensive automated way of calculating Flesch Reading Ease scores*" in Patient Disclosure Document 031350, US Patent Office,Washington, DC, 1974

[5] Smith, E. A. and R. J. Senter "*Automated readability index*" in AMRL-TR, 66-22 - Wright-Patterson AFB, OH: Aerospace Medical Division, 1967

[6] Coleman, M. and Liau, T.L. "*A computer readability formula designed for machine scoring*" in Journal of Applied Psychology, 60, No.2, 283-284, 1975

[7] Flesch R. "*A new readability yardstick*" in Journal of Applied Psychology, Vol. 32, pp. 221-233, 1948

[8] Flesch R. "*How To Test Readability*" in New York, Harper and Brothers, 1951

[9] Flesch R. "*The Art of Readable Writing*" in New York, Harper and Row., 1974

[10] Flesch, R. "*How to Write Plain English*", HarperCollins - 1st ed edition, August 1979

[11] Gunning, R. "*The Technique of Clear Writing*", McGraw-Hill New York, 1952

[12] H. Björnsson and Birgit Hård af Segerstad "*Lix på franska ochtio andra språk*", Stockholm,Pedagogiskt centrum, Stockholms skolförvaltning, 1979

[13] Anderson, J. "*The Challenges of the Multicultural Classroom*", Paper presented at the United Kingdom Reading Association Conference, Edinburgh, July 1981 (ED 207 023)

[14] Anderson, J. "*Analysing the Readability of English and Non-English Texts in the Classroom with Lix*", Paper presented at the Australian Reading Association Conference, Darwin, August 1981 (ED 207 022).

[15] Anderson J. "*A little-known readability index that is quick to use, reliable and easy to interpret*" in Journal of Reading, 1983

[16] McLaughlin, H "*SMOG grading - a new readability formula*" in Journal of Reading issue 22, pp639-646, 1969

[17] D. Berry, E. Kamsties, "*Ambiguity in requirements specification*", in Perspectives on Requirements Engineering, Kluwer 7-44, 2004

[18] Lami G., Gnesi S., Trentanni G., Fabbrini F., Fusani M. "*An Automatic Tool for the Analysis of Natural Language Requirements"* in Computer Systems Science and Engineering, vol. 20 (Vol.20, N. 1) pp. 53 - 62. CRL Publishing, 2005.

[19] Bucchiarone A., Fantechi A., Gnesi S., Lami G., Trentanni G. "*QuARS Express, an automatic analyzer of natural language requirements" i*n ASE 2008 - Automated Software Engineering, 2008. ASE 2008. 23rd IEEE/ACM International Conference on Automated Software Engineering (L'Aquila, Italy, 15-19 September 2008). Abstract, pp. 473 - 474. IEEE/ACM, 2008.

[20] Bucchiarone A., Fantechi A., Gnesi S., Trentanni G. "*An experience in using a tool for evaluating a large set of natural*

language requirements" in SAC '10 - 2010 ACM Symposium on Applied Computing (Sierre, Svizzera, 22-26 marzo 2010). Proceedings, pp. 281 - 286. Sung Y. Shin, Sascha Ossowski, Michael Schumacher, Michael Schumacher, Chih-Cheng Hung (eds.). ACM, 2010.

[21] QuARS (Quality Analyzer for Requirements Specification) http://www.quars.isti.cnr.it

[22] MODTRAIN: Innovative Modular Vehicle Concepts for an Integrated European Railway System. See: http://www.modtrain.com/.