Readability of Computer Systems Papers

Eitan Frachtenberg[1], Taylor Blair[2]

[1] Visiting Professor of Computer Science, Reed College
[2] Computer Science Student, Reed College

Introduction

Readability metrics are simple formulas in linguistics that return a number representing the difficulty level of a given piece of text. This number is an indicator of a reader's willingness to continue to engage with a given text (DuBay 2004). Readability is crucial because in order for a piece to be understood by a majority of adults in the United States, it needs to be written at the sixth grade level (Kirsch et al. 1993). Readability is integral to academic papers because an unreadable paper is more likely to be put down or ignored in spite of the high academic level of the reader.

Objectives

- Observe the readability gap between abstracts and fulltext.
- Remark on readability trends in subfields of computer system papers.
- Discuss the practicality of utilizing readability metrics to analyze computer system papers.

Methods

We have data from the previous summer on 2225 papers across 53 conferences. Over the course of this summer we converted PDFs of papers to TXT files and then hand cleaned the documents by removing artifacts, marking abstracts and placing the references on separate documents.

We initially did a finer, more methodical method of cleaning, but due to the time constraints we switched to a less rigorous hand cleaning method and relied instead on computerized cleaning.

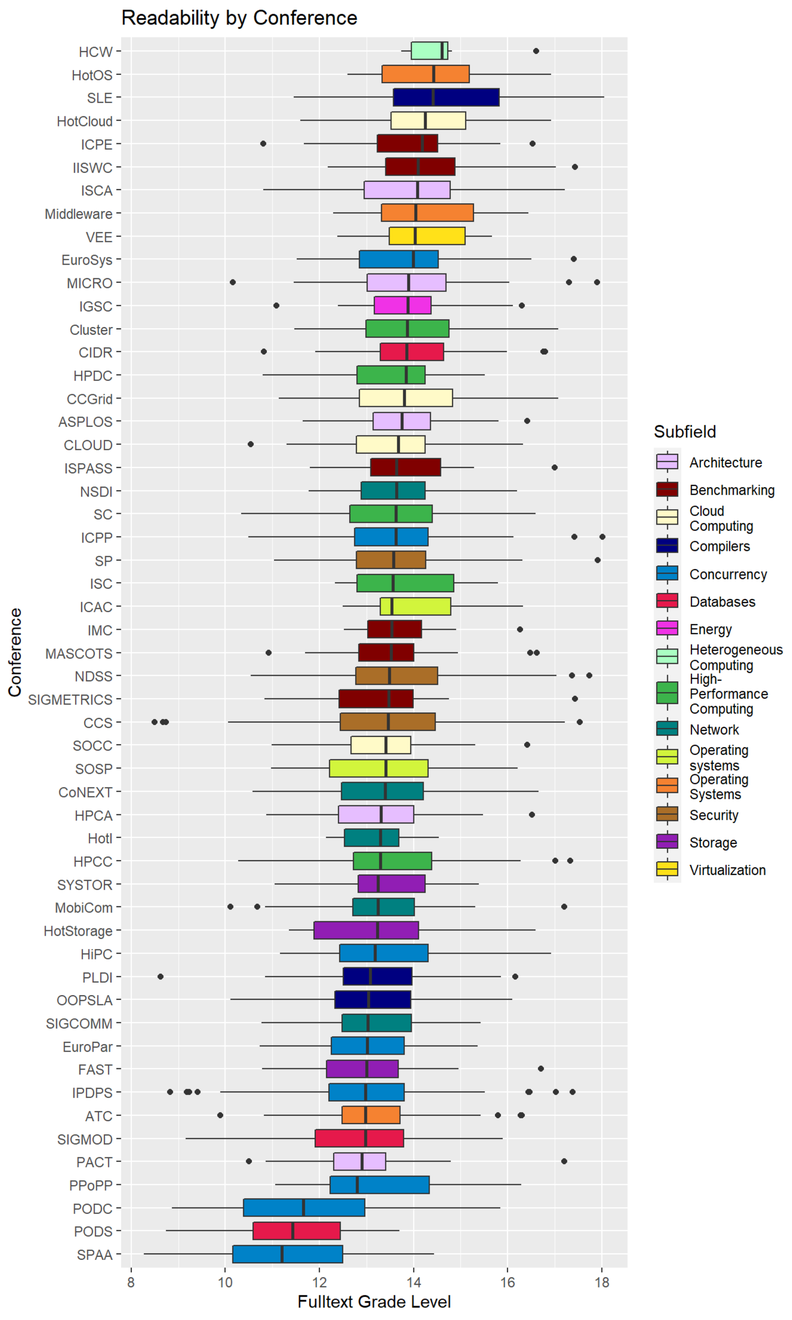
We then applied two readability formulas to our cleaned text files:

- **SMOG** (Simple Measure of Gobbledygook)
  - $\text{Grade Level} = 1.0430\sqrt{\text{Polysyllables} \times \frac{30}{\text{Sentences}}}$
  - A polysyllable is a word with three or more syllables.
  - This is a measure of the number of words with more than three syllables in a sentence.
- **Flesch–Kincaid**
  - $\text{Grade Level} = 206.835 - 1.015\left(\frac{\text{Words}}{\text{Sentences}}\right) - 84.6\left(\frac{\text{Syllables}}{\text{Words}}\right)$
  - Flesch–Kincaid compares the average number of words in a sentence to the average number of syllables in a word.

These formulas output a grade level (where lower is better) which allows us to compare texts.

Results

Previous summer research on computer systems scraped information about conference and paper topics. Using the data created this summer we merged these datasets to correlate between readability and topics.



The three most readable conferences, SPAA, PODS, and PODC were all hand-cleaned, removing all non-textual artifacts. The most readable papers ranked first due to the inline math. Inline math throws off readability metrics because the formula mistakes variables (i.e. x, y, z, θ) for one syllable words.

The three least readable conferences, SLE, HotOS, and HCW were cleaned using regex, with the exception of SLE, which was hand cleaned. These conferences papers often contained chunks of code (ex: `object Code extends StandardTokenParsers`) which skews the reading metric downward. This is because the readability formulas see the chunks as single sentences containing multiple polysyllables.

Table 1: Paper Subfield Readability

	Mean Readability Metric	
Paper Topic	**Abstract**	**Fulltext**
DB	16.157	12.799
Concurrency	16.155	13.046
Storage	15.810	13.091
Network	16.134	13.335
HPC	16.578	13.521
Compilers	16.121	13.537
Security	16.361	13.546
OS	16.205	13.566
Data	16.488	13.644
Architecture	16.382	13.649
GPGPU	16.452	13.720
Cloud	16.581	13.753
VM	16.576	13.765
Energy	16.496	13.912
Benchmark	16.800	13.970

On average, the abstract is 2.47 grades above the fulltext. Previous research indicated the same trend of abstracts being less readable compared to the fulltexts (Plavén-Sigray et al. 2017). One explanation is that when writing abstracts, authors are forced to write in a condensed space and use prose to compensate.

The most readable subfield by fulltext Flesch–Kincaid are database papers. The reason that database papers are given a highly readable grade is because the most common type of artifact was inline math.

In contrast, the least readable group, Benchmarks, contains large amounts of assembly code, functions and graphs containing text. These papers would have benefited from more hand cleaning.

Conclusion

While readability metrics have been useful in analyzing other fields, especially medical ones (Elhadad 2006), computer systems papers present a unique challenge with their embedded code, graphs and math equations. Through extensive data prep we are able to replicate findings about abstracts and fulltexts as well as some general trends about readability of papers by topic. The fact that the math-heavy and code-based occupy both extremes for readability demonstrates the need for a more labor-intensive hand cleaning.

References

DuBay, William H. 2004. "The Principles of Readability." *Online Submission*.

Elhadad, Noemie. 2006. "Comprehending Technical Texts: Predicting and Defining Unfamiliar Terms." In *AMIA Annual Symposium Proceedings*, 2006:239. American Medical Informatics Association.

Kirsch, Irwin S., Ann Jungeblut, Lynn Jenkins, and Andrew Kolstad. 1993. *Adult Literacy in America: A First Look at the Results of the National Adult Literacy Survey.* ERIC.

Plavén-Sigray, Pontus, Granville James Matheson, Björn Christian Schiffler, and William Hedley Thompson. 2017. "The Readability of Scientific Texts Is Decreasing over Time." *Elife* 6: e27725.