# What Happens when Laypersons Search Scientific Articles?

Jaap Kamps
University of Amsterdam

CLEF 2022 SimpleText Track, September 13, 2022, Bologna, Italy

# What Happens when Laypersons Search Scientific Articles?

Jaap Kamps
University of Amsterdam

CLEF 2022 SimpleText Track, September 13, 2022, Bologna, Italy

# Motivation

## Misinfo /Disinfo / Fake News

- Everyone agrees on the importance of **objective** and **reliable** information
- Citizens avoid scientific information as they assume it is **too complex**
- Can we better understand **barriers to access**? even remove them?

# Motivation

## Misinfo /Disinfo / Fake News

- Everyone agrees on the importance of **objective** and **reliable** information
- Citizens avoid scientific information as they assume it is **too complex**
- Can we better understand **barriers to access**? even remove them?

# What Happens When Laypersons Search Scientific Articles?

- Analysis of **Corpus** and **Popular Science**
  - *How Complex is Science?*
  - *Do Search Engines Use Complexity?*
- **Search** Experiments
  - *Can we Avoid Complexity?*
- **Text Simplification** Experiments
  - *Can we Simplify Scientific Text?*

# How Complex is Science?

Analysis #1

# Scientific Text Complexity

| **Grade Level** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **School** | *Elementary* | | | | | | *Jr. High* | | *High School* | | | | *Undergrad.* | | | | *Grad.* | *PhD* | | |
| | *Primary* | | | | | | *Secondary* | | | | | | *University* | | | | | *PhD* | | |
| | *Compulsory* | | | | | | | | | | | | *Higher Edu.* | | | | | | | |
| **Age** | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |

- Analyze Scientific abstracts, Popular science News articles, and Top 100 results
  - Using standard readability level measures (Flesch-Kincaid Grade Levels)

# Scientific Text Complexity

| **Grade Level** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **School** | *Elementary* | | | | | | *Jr. High* | | *High School* | | | | *Undergrad.* | | | | *Grad.* | *PhD* | | |
| | *Primary* | | | | | | *Secondary* | | | | | | *University* | | | | | *PhD* | | |
| | *Compulsory* | | | | | | | | | | | | *Higher Edu.* | | | | | | | |
| **Age** | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |

- Analyze Scientific abstracts, Popular science News articles, and Top 100 results
  - Using standard readability level measures (Flesch-Kincaid Grade Levels)
  - Target level is ~ 12 (high school diploma, exit compulsory education)

# Corpus, Context, and Requests

| Data | Sample Size | Length | | FKGL | |
|---|---|---|---|---|---|
| | | Mean | Median | Mean | Median |
| Corpus (scientific abstracts) | 8,513 | 951 | 905 | 14.55 | 14.40 |

- Corpus is too complex, corresponding to university level education

# Corpus, Context, and Requests

| Data | Sample Size | Length | | FKGL | |
|---|---|---|---|---|---|
| | | Mean | Median | Mean | Median |
| Corpus (scientific abstracts) | 8,513 | 951 | 905 | 14.55 | 14.40 |
| News (popular science) | 40 | 5,504 | 5,540 | 12.53 | 12.70 |

- Corpus is too complex, corresponding to university level education
- Popular science news is indeed the target level of 12!

# Corpus, Context, and Requests

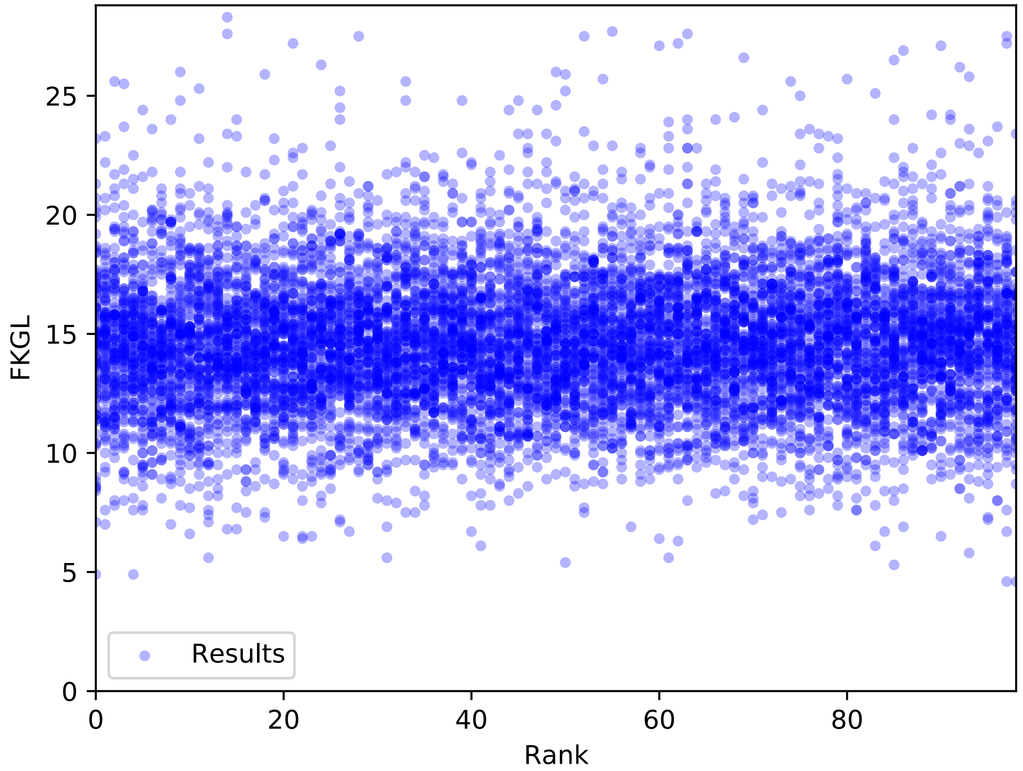
| Data | Sample Size | Length | | FKGL | |
|---|---|---|---|---|---|
| | | Mean | Median | Mean | Median |
| Corpus (scientific abstracts) | 8,513 | 951 | 905 | 14.55 | 14.40 |
| News (popular science) | 40 | 5,504 | 5,540 | 12.53 | 12.70 |
| Retrieved results (top 100) | 11,400 | 948 | 928 | 13.79 | 14.40 |

- Corpus is too complex, corresponding to university level education
- Popular science news is indeed the target level of 12!
- In response to a general query, the top 100 is as complex as the corpus...

# Do Search Engines Use Complexity?

Analysis #2

# Text Complexity per Rank of Retrieval




- There is no correlation between rank of retrieval and readability level!

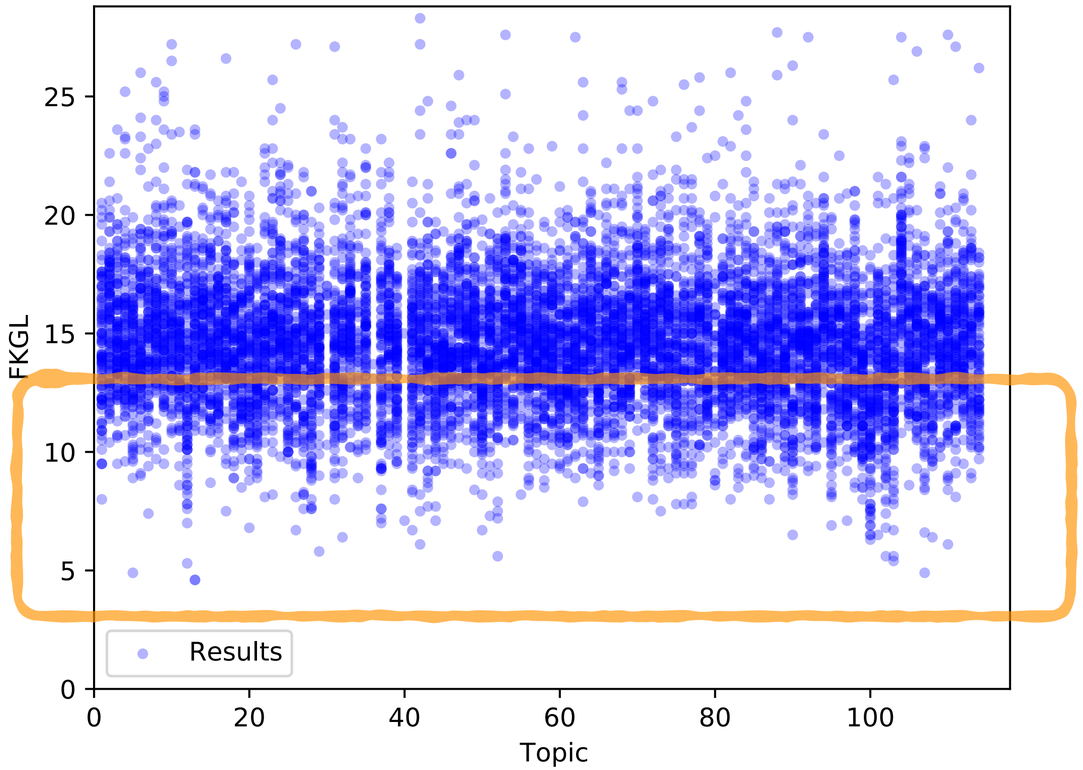
# #1 Scientific texts are too complex
# #2 Ranking ignores text complexity

Negative findings explaining why laypersons avoid science...

# Can we Avoid Complexity?

Strategy #1

# Complexity Variation per Topic




- For every request there are abstracts with the desirable readability level!

# Relevant+Simple: Complexity-Aware Ranking!

| Run | Top. | NDCG | | | FKGL | |
|---|---|---|---|---|---|---|
| | | 5 | 10 | 20 | Mean | Median |
| Elastic | 72 | 0.4053 | 0.4334 | 0.4438 | 13.79 | 14.40 |
| Automatic | 72 | 0.3531 | 0.3776 | 0.4073 | 11.70 | 12.80 |

- Per topic, we only keep the easier 1/2 of the abstracts retrieved by Elastic

# Relevant+Simple: Complexity-Aware Ranking!

| Run | Top. | NDCG | | | FKGL | |
|---|---|---|---|---|---|---|
| | | 5 | 10 | 20 | Mean | Median |
| Elastic | 72 | 0.4053 | 0.4334 | 0.4438 | 13.79 | 14.40 |
| Automatic | 72 | 0.3531 | 0.3776 | 0.4073 | 11.70 | 12.80 |

- Per topic, we only keep the easier 1/2 of the abstracts retrieved by Elastic
  - Small loss of precision (-13% NCDG@10)
    - Positive? avoids too complex (but judged relevant) abstracts!

# Relevant+Simple: Complexity-Aware Ranking!

| Run | Top. | NDCG | | | FKGL | |
|---|---|---|---|---|---|---|
| | | 5 | 10 | 20 | Mean | Median |
| Elastic | 72 | 0.4053 | 0.4334 | 0.4438 | 13.79 | 14.40 |
| Automatic | 72 | 0.3531 | 0.3776 | 0.4073 | 11.70 | 12.80 |

- Per topic, we only keep the easier 1/2 of the abstracts retrieved by Elastic
  - Small loss of precision (-13% NCDG@10)
    - Positive? avoids too complex (but judged relevant) abstracts!
  - **Relevant+Simple** leads to desired readability level of 12!

# #3 Per topic also readable abstracts
# #4 We can filter on readability levels

We can avoid abstracts with high text complexity!

# Can we Simplify Scientific Text?

Strategy #2

# Zero-shot Text Simplification

| Model | Task | Evaluated | SARI | Bleu | Precision | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | 1-gram | 2-gram | 3-gram | 4-gram |
| No change | Train | 648 | 0.5571 | 0.4204 | 0.6010 | 0.4531 | 0.3712 | 0.3089 |
| KiS Model | Train | 648 | 0.3984 | 0.2809 | 0.4881 | 0.3176 | 0.2319 | 0.1733 |

- Off-the-shelf Text Simplification model:
  - "Keep it Simple" (ACL/IJCNLP'21)
  - Used zero-shot, but can be trained *unsupervised* for scientific text
- Evaluation against human simplifications (train corpus)
  - BLEU 28% and SARI 40% (cmp. SARI on Wikipedia ~ 26-43%)

# Text Simplification: Readability Level

| Run | Task | Sentences | FKGL | | | Compression | |
|---|---|---|---|---|---|---|---|
| | | | Mean | Median | Ratio | Mean | Median |
| No change | Train | 648 | 15.46 | 15.4 | 0.00 | 1.00 | 1.00 |
| KiS Model | Train | 648 | 12.78 | 12.7 | 0.81 | 1.15 | 0.99 |
| No change | Test | 116,763 | 14.85 | 14.7 | 0.00 | 1.00 | 1.00 |
| KiS Model | Test | 116,763 | 12.06 | 11.9 | 0.79 | 1.33 | 1.01 |

- Evaluation must consider how much rewriting
  - No change has 15 FKGL (!)
  - Rewriting improves readability levels for 80% of the sentences
  - **Zero-shot model** leads to desired readability level of 12!

# Text Simplification: Examples

## 1 Results

### 1.1 G01.1

**2463945949** DIANE is a digital assistant system that aims to fasten allows the doctor a faster access to various informations at the patient and hospital such as health care facility , medical records , and also human resource data information . The fasten access This could be achieved by implementing done with face recognition and live streaming as part of the digital assistant system .

**2797641221** Digital assistants are emerging to become more prevalent becoming popular in our daily lives . In interacting with these assistants , It will allow users may engage in to do multiple tasks within in a short period of time faster way .

### 1.2 G01.2

**1448624402** As extensive experimental research has shown individuals Research showed that people suffer from diverse biases ( disproportionate weight in favor of or against an idea ) in decision-making . In our paper we analyze We analyzed the effects of decision-making biases of managers in collaborative decision processes on organizational performance . In the simulations , managerial managers decisions which are based on different levels of organizational complexity skills and different incentive systems rules , suffer from several known biases known from descriptive decision theory . The results illustrate how combined biases in combination with each other and in different organizational contexts affect organizational performance . We find that , contrary Contrary to intuition expectations , some combinations of biases significantly improve organizational performance while these biases negatively affect organizational performance . However , when they occur separately , they decrease performance . This might evoke considerations whether raises questions about the importance of rationality in decision-making should be as rational as possible .

### 1.3 G02.2

**2134216589** Big data , ( the enhanced ability to collect , store and analyze previously unimaginable huge quantities of data

**2943753439** The VUIs ( voice user interface ) like Amazon 's ' s Echo or Apple ' s Siri are popular nowadays . However , and Apple 's Siri they have drawn attention from different user groups ; however , these existing commercial VUIs support limited language options for users including native English speakers and non-native English speakers . It also shows that users ' English language proficiency 's skills play an important role in interacting with VUIs . The findings from results of this study can create insights for help VUI designers and developers for implementing while improving language options and more voice recognition algorithms in VUIs for different user groups across the world devices .

**2949291273** the The smart speakers can not distinguish differentiate between human voice from and machine voice . a A method to identify which of if the voice command is from a human or a machine is sending voice commands to a smart speaker is desired required . We propose a system composed of a speaker and microphones to detect the presence of humans . It could help to prevent such machine-voice based attacks to on a smart speaker with a machine voice in absence of residents , we propose a system consisting of a speaker and a microphone array to detect the existence of a human nearby , supposing it can be incorporated in a smart speaker in the future .

**2960901639** Although Even with the progress of speech recognition , and natural language processing ( NLP interactions between computers and human language ) have been greatly improved over the past few years , users may still may encounter get errors from time to time like " can not understand " , or " no requested audio resource ( such as music ) " , which can frustrate users . Therefore , when an error message is reported So , it is vital important that the smart speaker gives an effective and proper response during an error message . However The responses of the most popular brand of smart speakers are based on 2 elements , currently the response strategies adopted by leading smart speaker brands in China differed mainly on two dimensions : " apology or not " and " humor or neutral " . We explored studied user ' s preference of response strategies under in two error scenarios scenarios : " can not understand " and " no requested audio resource " . Two dependent variables ( satisfaction Satisfaction of the user and perceived sincerity of the response ) were measured . The results showed that participants users were more satisfied and perceived higher sincerity when smart speaker apologized in both error scenarios . In the " no requested audio resource " scenario , humor had no significant impact on the perception of satisfaction and or sincerity . But in the " can not understand " scenario , humorous expression humor decreased perceived sincerity .

## 1.5 G04.1

**2214798411** A pharmacophore analysis approach was used to investigate investigated and compare compared different classes compounds of compounds relevant to the drug discovery process ( specifically , drug molecules , compounds in high throughput screening libraries , combinatorial chemistry building blocks and nondrug molecules ) . Significant differences were observed between the pharmacophore profiles obtained for of the drug molecules and those obtained for the high-throughput screening compounds , which appear to be closely related to the nondrug pharmacophore distribution . It is suggested that the analysis of pharmacophore

# #5 Text simplification reduces complexity

We can reduce text complexity of scientific text!

# What Happens When Laypersons Search Scientific Articles?

#1 Scientific texts are too complex (FKGL 14-15)
#2 Ranking ignores text complexity

#3 Per topic also readable abstracts
#4 We can filter on readability levels (FKGL ~ 12)

#5 Text simplification reduces complexity (FKGL ~12)

# Q&A

Thanks to Femke Mostert, Mink Spronk, Ashmita Sampatsing, and David Rau!

More details in the paper http://ceur-ws.org/Vol-3180/paper-242.pdf

# Q&A

Thanks to Femke Mostert, Mink Spronk, Ashmita Sampatsing, and David Rau!

More details in the paper http://ceur-ws.org/Vol-3180/paper-242.pdf