



TURN IT SIMPLE



SimpleText Lab @CLEF-2022

Automatic Simplification of Scientific Text

Liana ERMAKOVA^{1,2}, Sílvia ARAÚJO³, Patrice BELLOT⁴, Radia HANNACHI^{5,2}, Stéphane HUET⁶, Jaap KAMPS⁷, Elise MATHURIN^{1,2}, Diana NURBAKOVA⁸, Irina OVCHINNIKOVA⁹, Eric SANJUAN⁶

1 - Université de Bretagne Occidentale, France
2 - HCTI - EA 4249, France
3 - University of Minho, Portugal
4 - Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France

5 - Université de Bretagne Sud, France
6 - Avignon Université, LIA, France
7 - University of Amsterdam, The Netherlands
8 - INSA Lyon, LIRIS UMR 5205 CNRS, France
9 - Sechenov University, Moscow, Russia



OBJECTIVE

To bring together researchers and practitioners working on the generation of simplified summaries of scientific texts

MOTIVATION

Scientific publications are difficult to read

Accessibility to:

- Non-native
- Younger readers
- Citizens with reading disabilities

Useful for:

- Scientific communication
- Science journalism
- Political communication
- Education

It has been observed that a domestic feline with a propensity to slothfulness and of an inherently obese nature was seated on the mat.

SIMPLIFICATION

The lazy fat cat sat on the mat.

Mike Unwalla. "Controlled language for text simplification". (2021) SimpleText-2021 Keynote

CHALLENGES AND STATE-OF-THE-ART

Content Selection

- Seq2Seq
- Text-to-Text Transformers

Complexity Spotting

Term recognition & extraction:

- IDF as term specificity
- WordNet distance to the basic terms
- Text-to-Text Transformers

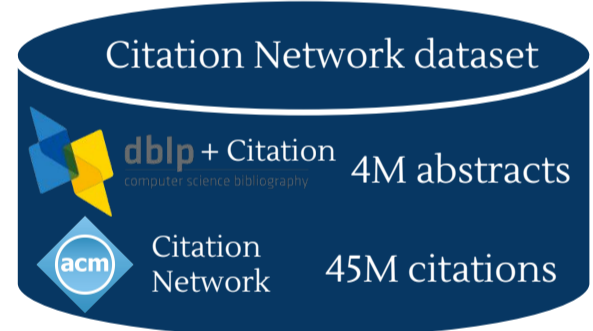
Text Simplification

Rule-based methods → Statistical Machine Translation

- Seq2Seq
- Text-to-Text Transformers

Y. Liu, M. Lapata. "Text Summarization with Pretrained Encoders." arXiv:1908.08345 [cs] (Sep 2019)
A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever. "Language Models are Unsupervised Multitask Learners" O. Lieber, O. Sharir, B. Lentz, Y. Shoham. "Jurassic-1: Technical Details and Evaluation"
L. Xue, N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, C. Raffel. "mT5: A massively multilingual pre-trained text-to-text transformer" (ACL 2021)
Chao Jiang, Mounica Madhala, Wuwei Lan, Yang Zhong, Wei Xu. "Neural CRF Model for Sentence Alignment in Text Simplification" (ACL 2020)

DATA COLLECTION



Around Computer science



Digital assistants like Siri and Alexa entrench gender biases, says UN x13

Topic 1: Digital assistants

Topic 2: Biases

<https://www.theguardian.com/technology/2019/may/22/digital-voice-assistants-siri-alexa-gender-biases-unesco-says>

TASK 1: What is in (or out)?

Select passages to include in a simplified summary, given a query

Topic 1: Digital assistants

1

People are becoming increasingly comfortable using Digital Assistants (DAs) to interact with services or connected objects.

2

Automated decision making based on big data and machine learning (ML) algorithms can result in discriminatory decisions against certain protected groups defined upon personal data like gender, race, sexual orientation etc.

3

Such algorithms designed to discover patterns in big data might not only pick up any encoded societal biases in the training data, but even worse, they might reinforce such biases resulting in more severe discrimination.

TASK 3: Rewrite this!

Given a query, simplify passages from scientific abstracts

Source passage

Automated decision making based on big data and machine learning (ML) algorithms can result in discriminatory decisions against certain protected groups defined upon personal data like gender, race, sexual orientation etc. Such algorithms designed to discover patterns in big data might not only pick up any encoded societal biases in the training data, but even worse, they might reinforce such biases resulting in more severe discrimination.

Simplified passage

Automated decision-making may include sexist and racist biases and even reinforce them because their algorithms are based on the most prominent social representation in the data collection they use.

TASK 2: What is unclear?

Given a passage and a query, rank terms/concepts that are required to be explained for understanding this passage (definitions, context, applications,..)

Source passage

Automated decision making based on big data and machine learning (ML) algorithms can result in discriminatory decisions against certain protected groups defined upon personal data like gender, race, sexual orientation etc. Such algorithms designed to discover patterns in big data might not only pick up any encoded societal biases in the training data, but even worse, they might reinforce such biases resulting in more severe discrimination.

1 Machine learning

2 Societal biases

3 ML

OPEN TASK

Found any other fascinating use of our data? Feel free to share with us!