Introduction
○○○

SimpleText 2022
○○○

SimpleText 2023
○○○

CLEF 2023 SimpleText Track:
Automatic Simplification of Scientific Texts

Liana Ermakova *et al.*

CLEF - **September 8, 2022**

Motivation

- Scientific documents are difficult to understand
- Accessibility to:
    - Non-native
    - Younger readers
    - Citizens with reading disabilities
- Useful for:
    - Scientific communication
    - Science journalism
    - Political communication
    - Education

Goals

- To create a simplified summary of multiple scientific documents based on a query which provides users with an instant simplified overview on the specific topic they are interested in

---

- **Technical & evaluation** challenges of scientific text simplification
- To provide appropriate reusable **data** and **benchmarks** for text simplification

Introduction
○○●

SimpleText 2022
○○○

SimpleText 2023
○○○

Example (zero-shot text simplification)

- *Scientific Abstract (FKGL 17.0 – University grad. school)*
  Searching scientific literature and understanding technical scientific documents can be very difficult for users as there are a vast number of scientific publications on almost any topic and the language of science, by its very nature, can be complex. Scientific content providers and publishers should have mechanisms to help users with both searching the content in an effective way and understanding the complex nature of scientific concepts. . . .

- *GPT-2 revisions (FKGL 12.9 – High school diploma)*
  Searching for scientific literature ~~and understanding technical scientific documents~~ can be very ~~difficult~~ time-consuming for users as there are a vast number of scientific publications on almost any topic and the language of science , by its very nature , can be ~~complex~~ very confusing . Scientific content providers and publishers should have mechanisms to help users ~~with both searching~~ find the ~~content~~ right information in an effective way , and understanding the ~~complex~~ nature of scientific concepts . . . .

Introduction
000

SimpleText 2022
●00

SimpleText 2023
000

SimpleText Track at CLEF 2022

Shared tasks = pipeline:

- *Task 1: What is in (or out)?*
  - Select passages to include in a simplified summary, given a query.
- *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,..).
- *Task 3: Rewrite this!*
  - Given a query, simplify passages from scientific abstracts.

Introduction
ooo

SimpleText 2022
o●o

SimpleText 2023
ooo

Corpus & Topics 2022

- Citation Network Dataset: DBLP+Citation, ACM Citation network
  - 4,232,520 abstracts in English
- Topics = 40 press articles + manually extracted queries (keywords)
  - 20 articles from *The Guardian*
  - 20 articles from *Tech Xplore*
- Train data for **Task 3**: parallel corpus of 648 manually simplified sentences
- Test data for **Tasks 2 & 3**: 116,763 sentences from the DBLP abstracts according to the queries from Task 1

Introduction
○○○

SimpleText 2022
○○●

SimpleText 2023
○○○

SimpleText 2022 Statistics

- 62 registered teams
- 40 users downloaded data from the server

Team	Task 1	Task 2	Task 3	Total runs
UAms	2	1		3
NLP@IISERB	3			3
SimpleScientificText		1		1
aaac		1		1
LEA_T5		1	1	2
PortLinguE			1	1
CYUT Team2	1		1	2
HULAT-UC3M			10	10
CLARA-HD			1	1
Total runs	6	4	14	24

Introduction
000

SimpleText 2022
000

SimpleText 2023
●00

Shared Tasks 2023

- *Task 1: Content Selection: Selecting passages to include in a simplified summary*
  - topical relevance
  - $+$ text complexity
  - $+$ source authoritativeness
- *Task 2: Complexity Spotting: Identifying and explaining difficult concepts for laypersons*
  - difficult term detection
  - $+$ usefulness of the provided explanation with regard to a query
  - $+$ difficulty of the provided explanation
- *Task 3: Text Simplification: Scientific text simplification*
  - manual evaluation of information distortion & text complexity
  - $+$ expand the training and evaluation data
  - $+$ large-scale automatic evaluation measures

Introduction
○○○

SimpleText 2022
○○○

SimpleText 2023
○●○

Other News

- Elsevier has become SimpleText partner
- New organizers:
  - Olivier Augereau, ENIB, Lab-STICC CNRS UMR 6285, France
  - Hosein Azarbonyad, Elsevier, The Netherlands (to be confirmed)
  - Giorgio Maria Di Nunzio, Università di Padova, Italy (to be confirmed)
- **A PhD student to be hired!**

Introduction
○○○

SimpleText 2022
○○○○

SimpleText 2023
○○●

Thank you !
Participate in our track !

Website : https://simpletext-project.com
E-mail : contact@simpletext-project.com
Twitter : https://twitter.com/SimpletextW
Google group : https://groups.google.com/g/simpletext