Text Simplification of Scientific Texts for Non-Expert Readers

SimpleText@CLEF 2023
Task 3: Rewrite this!

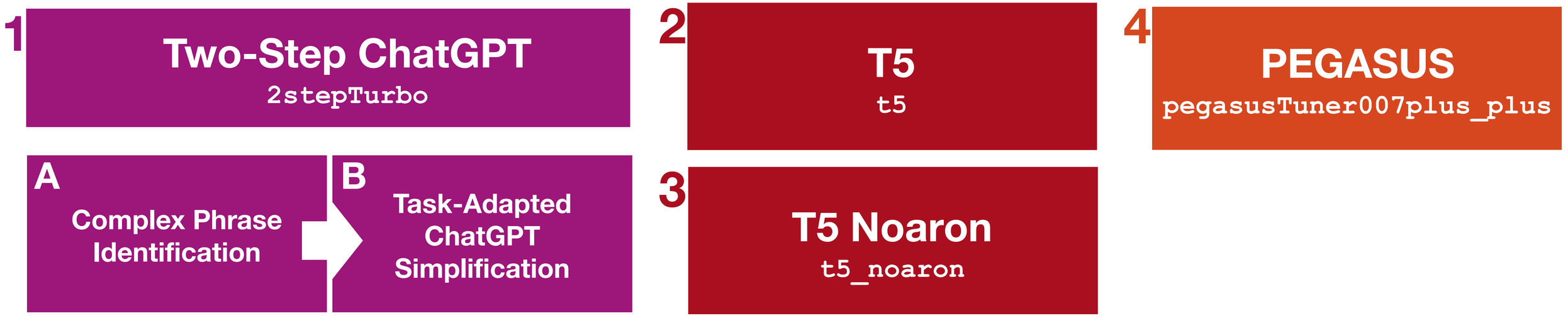
Team irgc

Björn Engelmann, **Fabian Haak**, Christin Katharina Kreutz, Narjes Nikzad Khasmakhi, Philipp Schaer

September 19th 2023

Technology
Arts Sciences
TH Köln

Overview of our Runs for Task 3: Text Simplification - Rewriting Scientific Text



Björn Engelmann, **Fabian Haak**, Christin Katharina Kreutz, Narjes Nikzad Khasmakhi, Philipp Schaer

Team irgc – Information Retrieval Research Group at TH Köln

Technology Arts Sciences TH Köln

Motivation and Goals

SARI

ROUGE

compression

readability

?!

Evaluation Criteria of SimpleText

false or unsupported information

style

overgeneralization

contradictions

unnecessary details

ambiguity

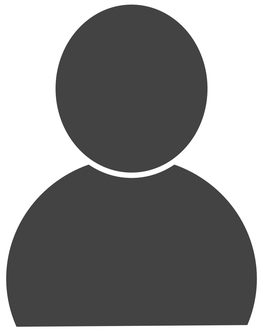
omission of essential information

redundancy

oversimplification

topic shift

nonsense



Non-Expert Reader

abbreviations

complex scientific phrases

typical scientific syntax

Björn Engelmann, **Fabian Haak**, Christin Katharina Kreutz, Narjes Nikzad Khasmakhi, Philipp Schaer

Team irgc – Information Retrieval Research Group at TH Köln

Technology Arts Sciences TH Köln

Run: ChatGPT “2stepTurbo”

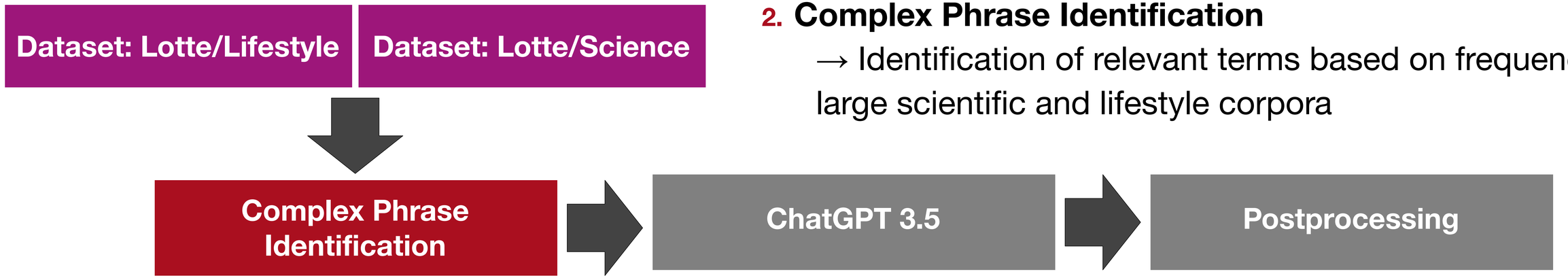
Björn Engelmann, **Fabian Haak**, Christin Katharina Kreutz,
Narjes Nikzad Khasmakhi, Philipp Schaer

Team irgc – Information Retrieval Research Group at TH Köln

Technology
Arts Sciences
TH Köln

Run: ChatGPT “2stepTurbo” – Complex Phrase Identification

1. **Phrase Identification**
   → Identification of n-gram keyphrases via 🤗 model (KBIR-inspec)
2. **Complex Phrase Identification**
   → Identification of relevant terms based on frequencies in large scientific and lifestyle corpora



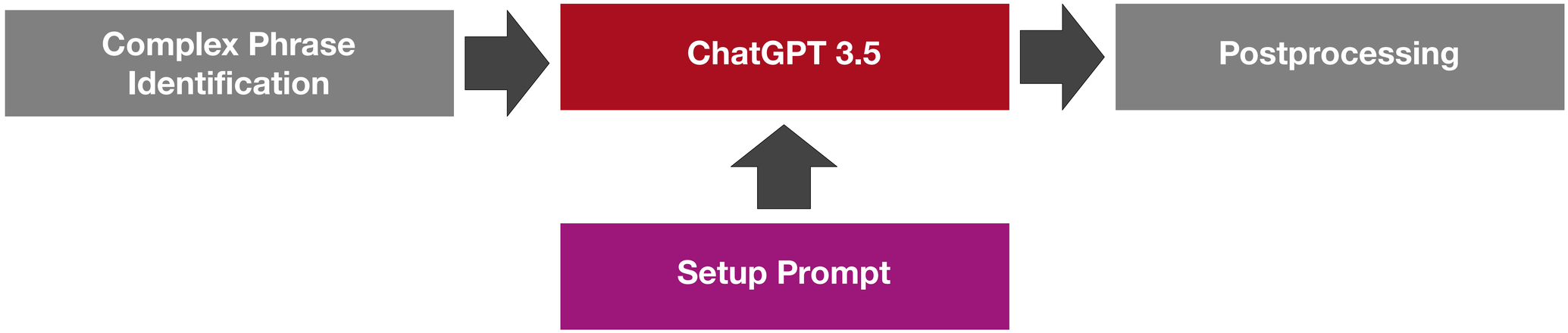
Indeed, modelling of an [infection network] between viral and cellular proteins will provide a conceptual and analytic framework to efficiently formulate new [biological hypothesis] at the [proteome] scale and to rationalize [drug discovery].

Björn Engelmann, **Fabian Haak**, Christin Katharina Kreutz,
Narjes Nikzad Khasmakhi, Philipp Schaer

Team irgc – Information Retrieval Research Group at TH Köln

Technology
Arts Sciences
TH Köln

Run: ChatGPT “2stepTurbo” – ChatGPT



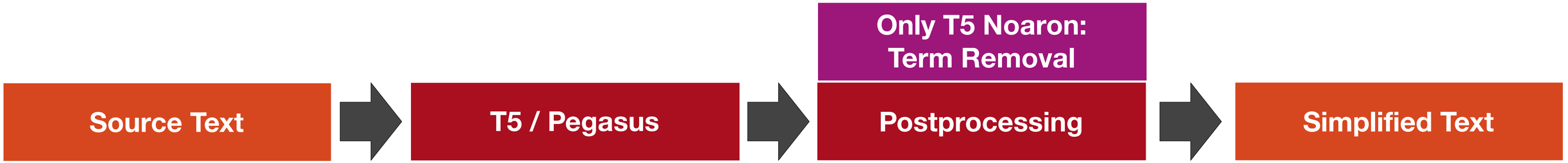
- Assessed: chain-of-though, few-shot, zero-shot
- Setup prompt:
  - Task is to **simplify for non-expert readers**
  - **Complex phrases indicated** in text
    → Need to be simplified or explained

Björn Engelmann, **Fabian Haak**, Christin Katharina Kreutz,
Narjes Nikzad Khasmakhi, Philipp Schaer
Team irgc – Information Retrieval Research Group at TH Köln
Technology Arts Sciences TH Köln

Runs: Summarization with T5 and PEGASUS

- Pretrained T5-base and PEGASUS by Tuner007 via 🤗
- For T5 Noaron: Removal of frequently AI hallucination artefact:

  aaron carroll: modelling of an infection network between viral and cellular proteins will provide a conceptual framework.



Björn Engelmann, **Fabian Haak**, Christin Katharina Kreutz,
Narjes Nikzad Khasmakhi, Philipp Schaer

Team irgc – Information Retrieval Research Group at TH Köln

Technology
Arts Sciences
TH Köln

Evaluation and Run Prioritization

- Focus on manual evaluation
- T5_noaron best performing model according to quant. measures
- Manual evaluation: ChatGPT approach **least nonsense and hallucination, most meaning preservation, best abbreviation handling**

measure	source texts	ChatGPT	t5	t5_noaron	pegasus
empty texts ↓	-	3	148	148	**1**
compression ↑	-	1.16 ± 0.72	1.53 ± 2.3	**1.55** ± 2.34	1.0 ± 1.6
Flesch ↑	36.7 ± 23.33	42.41 ± 21.61	**46.4** ± 38.65	46.3 ± 38.85	38.79 ± 23.28
Dale-Chall ↓	12.68 ± 2.18	12.35 ± 1.96	12.41 ± 3.05	12.37 ± 3.05	**11.44** ± 2.62
difficult words ↓	8.51 ± 4.32	7.55 ± 3.33	6.23 ± 2.86	**6.2** ± 2.88	7.99 ± 3.83
reading time ↓	1.94 ± 0.99	1.71 ± 0.68	1.45 ± 0.59	**1.44** ± 0.59	2.24 ± 1.03
syllable count ↓	39.12 ± 20.07	34.4 ± 13.81	29.22 ± 12.22	**29.11** ± 12.25	45.23 ± 20.86
lexicon count ↓	22.51 ± 11.05	20.26 ± 8.04	16.88 ± 7.04	**16.8** ± 7.07	26.18 ± 11.73
sentence count ↓	**1.05** ± 0.58	1.07 ± 0.29	1.44 ± 0.56	1.44 ± 0.56	1.32 ± 0.62

Björn Engelmann, **Fabian Haak**, Christin Katharina Kreutz,
Narjes Nikzad Khasmakhi, Philipp Schaer
Team irgc – Information Retrieval Research Group at TH Köln

Evaluation and Run Prioritization: Manual Evaluation

Source Sentence	ChatGPT “2stepTurbo”
There are two challenges in analyzing the NHPu0027s surveillance video: the NHPu0027s behaviors are lack of regularity and intention, and serious occlusions are brought by the fences of the cages.	The irregular behavior of non-human primates and occlusions caused by cages' fences make it challenging to analyze their surveillance video footage.

Most noteworthy:

- Replacement of abbreviations
- Resolution of unicode formatting issues
- Resolution of syntactical-deductive sentence structure
- Other linguistic simplification (vocabulary, sentence length, reformulation)

Björn Engelmann, **Fabian Haak**, Christin Katharina Kreutz,
Narjes Nikzad Khasmakhi, Philipp Schaer

Team irgc – Information Retrieval Research Group at TH Köln

Technology
Arts Sciences
TH Köln

Conclusion

- 4 submitted runs
- Motivation: Focus on target-audience-specific aspects
  - Priority based on manual evaluation
  - Criteria from task definition and derived from dataset
- ChatGPT with complex phrase identification best according to manual evaluation, performed well in Lab (**SARI 46.98**)
- T5 and PEGASUS best according to quantitative metrics, but perform worse in SARI (T5: 37.83, T5_noaron: 37.84, PEGASUS: 23.28)

Björn Engelmann, **Fabian Haak**, Christin Katharina Kreutz,
Narjes Nikzad Khasmakhi, Philipp Schaer

Team irgc – Information Retrieval Research Group at TH Köln

Technology
Arts Sciences
TH Köln