Overview of SimpleText 2021 - CLEF Workshop on Text Simplification for Scientific Information Access

Liana Ermakova[1(✉)], Patrice Bellot[2], Pavel Braslavski[3], Jaap Kamps[4], Josiane Mothe[5], Diana Nurbakova[6], Irina Ovchinnikova[7], and Eric SanJuan[8]

[1] Université de Bretagne Occidentale, HCTI - EA 4249, Brest, France
liana.ermakova@univ-brest.fr
[2] Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France
[3] Ural Federal University, Yekaterinburg, Russia
[4] University of Amsterdam, Amsterdam, The Netherlands
[5] Université de Toulouse, IRIT, Toulouse, France
[6] Institut National des Sciences Appliquées de Lyon, Lyon, France
[7] Sechenov University, Moscow, Russia
[8] Avignon Université, LIA, Avignon, France

Abstract. Information retrieval has moved from traditional document retrieval in which search is an isolated activity, to modern information access where search and the use of the information are fully integrated. But non-experts tend to avoid authoritative primary sources such as scientific literature due to their complex language, internal vernacular, or lacking prior background knowledge. Text simplification approaches can remove some of these barriers, thereby avoiding that users rely on shallow information in sources prioritizing commercial or political incentives rather than the correctness and informational value. The CLEF 2021 SimpleText track addresses the opportunities and challenges of text simplification approaches to improve scientific information access head-on. We aim to provide appropriate data and benchmarks, starting with pilot tasks in 2021, and create a community of NLP and IR researchers working together to resolve one of the greatest challenges of today.

Keywords: Scientific text simplification · (Multi-document) summarization · Contextualization · Background knowledge

> Everything should be made as simple as possible, but no simpler
>
> *Albert Einstein*

1 Introduction

Scientific literacy, including health related questions, is important for people to make right decisions, evaluate the information quality, maintain physiological

© Springer Nature Switzerland AG 2021
K. S. Candan et al. (Eds.): CLEF 2021, LNCS 12880, pp. 432–449, 2021.
https://doi.org/10.1007/978-3-030-85251-1_27

and mental health, avoid spending money on useless items. For example, the stories the individuals find credible can determine their response to the COVID-19 pandemic, including the application of social distancing, using dangerous fake medical treatments, or hoarding. Unfortunately, stories in social media are easier for lay people to understand than the research papers. Scientific texts such as scientific publications can also be difficult to understand for non domain-experts or scientists outside the publication domain. Improving text comprehensibility and its adaptation to different audience remains an unresolved problem. Although there are some attempts to tackle the issue of text comprehensibility, they are mainly based on readability formulas, which have not convincingly demonstrated the ability to reduce the difficulty of text [30].

To put a step forward to automatically reduce difficulty of text understanding, we propose a new workshop called SimpleText which aims to create a community interested in generating simplified summaries of scientific documents. Thus, the goal of this workshop is to connect researchers from different domains, such as Natural Language Processing, Information Retrieval, Linguistics, Scientific Journalism etc. in order to work together on automatic popularization of science.

Improving text comprehensibility and its adaptation to different audience bring societal, technical, and evaluation challenges. There is a large range of important *societal challenges* SimpleText is linked to. Open science is one of them. Making the research really open and accessible for everyone implies providing it in a form that can be readable and understandable; referring to the "comprehensibility" of the research results, making science understandable [20]. Another example of those societal challenges is offering means to develop counter-speech to fake news based on scientific results. SimpleText also tackles *technical challenges* related to data (passage) selection and summarization, comprehensibility and readability of texts.

To face these challenges, SimpleText provides an open forum aiming at answering questions like:

- **Information selection:** Which information should be simplified (e.g., in terms of document and passage selection and summarisation)?
- **Comprehensibility:** What kind of background information should be provided (e.g., which terms should be contextualized by giving a definition and/or application)? What information is the most relevant or helpful?
- **Readability:** How to improve the readability of a given short text (e.g., by reducing vocabulary and syntactic complexity) without information distortion?

We provides data and benchmarks, and addresses evaluation challenges underlying the technical challenges, including:

- How to evaluate information selection?
- How to evaluate background information?
- How to measure text simplification?

2 Related Work

In order to simplify scientific texts, one has to (1) select the information to be included in a simplified summary, (2) decide whether the selected information is sufficient and comprehensible or provide some background knowledge if not, (3) improve the readability of the text [15]. Our workshop is organized around this pipeline.

2.1 Information Selection

People have to manage the constantly growing amount of information, e.g. according to research platform Dimensions[1], from 01/01/20–01/10/20, about 180K articles on COVID-19 were published. To deal with this data volume, a concise overview, i.e. a summary, is needed. Thus, summarization is already a step towards text simplification as it reduces the amount of information to be processed. Besides, people prefer to read a short document instead of a long one. Since motivation to understand a scientific text is of importance for readers, the simplified options depends on the motivation of readers [38]. Thus, the information in a summary designed for a scientist from a specific field should be different from that adapted for general public and we should take into account differences in narrative and information texts comprehension while evaluating the comprehensibility level of simplified texts in different readership. Thus, the main challenge is to choose which information should be included in a simplified text. Despite recent significant progress in the domains of information retrieval (IR) and natural language processing (NLP), the problem of constructing a consistent overview has not been solved yet [17].

Automatic summarization can simplify access to primary source scientific documents - the resulting concise text is expected to highlight the most important parts of the document and thus reduces the reader's efforts. Evaluation initiatives in the 2000s such as the Document Understanding Conference (DUC) and the Summarization track at the Text Analysis Conference[2] (TAC) have focused primarily on the automatic summarization of news in various contexts and scenarios. Scientific articles are typically provided with a short abstract written by the authors. Thus, automatic generation of an abstract for a standalone article does not seem to be a practical task. However, if we consider a large collection of scientific articles and citations between them, we can come to the task of producing an abstract that would contain the important aspects of a paper from the perspective of the community. Such a task has been offered to the participants of the TAC 2014 Biomedical Summarization Track, as well as of the CL-SciSumm shared task series. Another close work is CLEF-IP 2012–2013: Retrieval in the Intellectual Property Domain (novelty search). Given a patent claim, the task was to retrieve the passages relevant to this claim from a document collection; the retrieved passages were compared to the relevant passages indicated by a patent examiner in her/his search report, but this relevancy

[1] https://www.dimensions.ai.
[2] https://tac.nist.gov/2014/BiomedSumm.

relationship between claims and text passages in other documents cannot be considered as text simplification nor summarization.

Sentence selection is a crucial but understudied task in document simplification [59] as existing works mainly focus on word/phrase-level (simplification of difficult words and constructions) [5,23,34,44,49,57] or sentence-level simplifications [9,14,51,56,60,61]. The state-of-the-art in automatic summarization is achieved by deep learning models, in particular by pretrained Bidirectional Encoder Representations from Transformers (BERT) which can be used for both extractive and abstractive models [32] However, the information in a summary designed for an expert might be different from that for a general audience. Therefore, a major step in training artificial intelligence (AI) text simplification models is the creation of high quality data. Zhong et al. studied various discourse factors associated with sentence deletion on the Newsela corpus containing manually simplified sentences from news articles [59] (contrary to SimpleText which focuses on scientific literature). They found that professional editors utilize different strategies to meet the readability standards of elementary and middle schools. It is important to study the limits of existing models, like GPT-2 for English and CamemBERT for French [35], and how it is possible to overcome them.

How to evaluate the information in a simplified summary? Summary informativeness metrics can mainly be divided into two classes: (1) questionnaire-based metrics and (2) overlap-based metrics [17]. In case of questionnaire-based metrics, an assessor should answer a set of questions issued from the source text or evaluate the importance of each sentence/passage [17], e.g. Responsiveness metric was introduced at the Document Understanding Conference (DUC) [42]. A Pyramid score is in the middle between the questionnaire based and overlap-based metrics since it calculates the number of repetitions of information units of variable length inside a sentence labeled by experts in their own words [41]. Overlap-based measures estimate the proportion of shared words between the reference summary and the summary under consideration, e.g. a widely used ROUGE metric (short for Recall-Oriented Understudy for Gisting Evaluation) and its variants [31]. The overlap metrics require a set of reference summaries. Providing a collection of simplified texts makes it possible to apply overlap metrics like ROUGE to text simplification.

2.2 Comprehensibility (Background Knowledge)

Comprehensibility of a text varies for different readerships. Readers of popular science texts have a basic background, are able to process logical connections and recognize novelty [26]. In the popular science text, a reader looks for rationalization and clear links between well known and new [39]. In order to really understand new concepts, readers need to include them into their mental representation of the scientific domain. Models of mental representation of knowledge are mostly based on propositional structures, but we consider embodied (grounded) reading comprehension to be useful for the SimpleText project because embodied cognition can provide a mental bridge between a personal experience and semantic representation of knowledge in the long-term semantic memory [47]. Therefore, a simplified scientific text has to be able to evoke clear associations with embodied cognition.

According to The Free Dictionary[3], background knowledge is "information that is essential to understanding a situation or problem". The lack of basic knowledge can become a barrier to reading comprehension and there is a knowledge threshold allowing reading comprehension [43]. Scientific text simplification presupposes the facilitation of readers' understanding of complex content by establishing links to basic lexicon, avoiding distortion connections among objects within the domain. Traditional methods of text simplification try to eliminate complex concepts and constructions [5,23,34,44,49,57]. However, it is not always possible, especially in the case of scientific literature. In contrast to previous research, SimpleText is not limited to a "Split and Rephrase" task but also aims to provide a sufficient context to a scientific text as the lack of background knowledge could be a major obstacle for text comprehension [43]. Entity linking (Wikification, task of tying named entities from the text to the corresponding knowledge base items, e.g. Wikipedia) could help mitigate the background knowledge problem, by providing definitions, illustrations, examples, and related entities. However, the existing entity linking datasets are focused primarily on such entities as people, places, and organizations [25], while a lay reader of a scientific article needs assistance with new concepts and methods. Wikification is close to the task of terminology and key-phrase extraction from scientific texts [3]. The idea of contextualizing news was further developed in the Background Linking task at TREC 2020 News Track aiming at a list of links to the articles that a person should read next [2]. It is also important to remember that the goal is to keep the text simple and short, as long texts can discourage potential readers. Thus, in contrast to previous projects, SimpleText aims to provide lacking background knowledge but keeping the text as short as possible in order to help a user understand a complex text which cannot be further simplified without severe information distortion. Searching for background knowledge is close to INEX/CLEF Tweet Contextualization track 2011–2014 [4] and CLEF Cultural micro-blog Contextualization 2016, 2017 Workshop [18], but SimpleText differs from them by making a focus on a selection of notions to be explained and the helpfulness of the information provided rather than its relevance.

2.3 Readability (Language Simplification)

Sentence compression can be seen as a middle ground between text simplification and summarization. The task is to remove redundant or less important parts of an input sentence, preserving its grammaticality and original meaning. Recent works have applied the BERT neural network model [19,36,58], in order to simplify sentences. These approaches are mainly reduced to the "Split and Rephrase" task. Moreover, simplification systems are mainly limited by deleting

[3] https://www.thefreedictionary.com/background+knowledge.

words [33]. Besides, although large pre-trained BERT models like GPT2 outperformed other state-of-the-art models on several NLP tasks, researchers point to several serious issues of these models – consistency and coherency (coreference errors) [52]. In any case, to train and evaluate an AI model one should have a corpus of scientific articles and their simplified versions with a benchmarking system. In previous works, some datasets were developed such as WebSplit et WikiSplit, however the text simplification task was reduced to "Split and Rephrase" [1,6,40]. Another dataset was based on Simple Wikipedia but there is no direct correspondence between Wikipedia and Simple Wikipedia articles [11]. The comparable WikiLarge dataset combines aligned sentence pairs in [29], the aligned and revision sentence pairs in [53], and WikiSmall corpus [60]. To have parallel data (not comparable) is important as the efficiency of a text simplification system depends on the quality and quantity of training data [27]. The dataset Newsela contains 1,932 English news articles re-written by professional editors into four simpler versions [55]. In contrast to that, we focus on scientific texts. CL-SciSumm-2020 features LaySummary subtask[4], where a participating system must produce a text summary of a scientific paper (overall scope, goal and potential impact without using technical jargon) on epilepsy, archaeology, and materials engineering intended for a non-technical audience. However, in most cases, the names of the objects are not replaceable in the process of text transformation or simplification due to the risk of information distortion [12,37]. In this case, complex concepts should be explained to a reader.

Grabar and Cardon introduced a corpus of technical and simplified medical texts in French [7,24]. The corpus contains 663 pairs of comparable sentences issued from encyclopedias, drug leaflets and scientific summaries, and aligned by two annotators. In [7], they proposed an automatic method for sentence alignment. In their further work, using different ratios of general and specialized sentences, they trained neural models on (1) the health comparable corpus in French, (2) the WikiLarge corpus translated from English to French, and (3) and a lexicon that associates medical terms with paraphrases [8]. Jiang et al. proposed a neural CRF alignment model and constructed two text simplification datasets: Newsela-Auto and Wiki-Auto [27]. Their transformer-based seq2seq model established a new state-of-the-art for text simplification in both automatic and human evaluation. In contrast to that, our corpus is not comparable (when simplified sentences are not issued from original sentences but are similar to them), but parallel (source sentences are directly simplified, so they carry the same information). Besides, their work tackles language simplification only without considering content selection for popularized texts which can be different from those designed for experts.

Readability formulas have not convincingly demonstrated the ability to reduce the difficulty of the text [10,21,30,48]. Automatic evaluation metrics have been designed to measure the results of text simplification: SARI [55] targets lexical complexity, while SAMSA estimates the structural complexity of a sentence [50]. Formality style transfer is a cognate task, where a system rewrites

[4] https://ornlcda.github.io/SDProc/sharedtasks.html#laysumm.

a text in a different style preserving its meaning [46]. These tasks are frequently evaluated with lexical overlap metrics such as BLEU [45] or ROUGE [31] to compare the system's output against gold standard. SimpleText is also aimed at providing adequate evaluation metrics for text simplification. Since traditional readability indices can be misleading [54], we rely on human evaluation.

3 Data Set

3.1 Collection

For this edition we use the Citation Network Dataset: DBLP+Citation, ACM Citation network[5]. An elastic search index is provided to participants accessible through a GUI API. This Index is adequate to:

– apply basic passage retrieval methods based on vector or language IR models;
– generate Latent Dirichlet Allocation models;
– train Graph Neural Networks for citation recommendation as carried out in StellarGraph[6] for example;
– apply deep bi directional transformers for query expansion;
– and much more ...

While structured abstracts with distinct, labeled sections for rapid comprehension are an emerging trend since they tend to be informative [16,22], several approaches were proposed to classify sentences in non-structured abstracts [13,16,28]. However, non-expert are usually interested in other types of information. We selected passages that are adequate to be inserted as plain citations in the original journalistic article. The comparison of the journalistic articles with the scientific ones as well as the analysis we carried out to choose topics demonstrated that non-expert, the most important information is the application of an object (which problem can be solved? how to use this information/object? what are examples?).

One of the important problems in manual text simplification is a cognitive bias called the curse of knowledge, which occurs when an individual assumes that their interlocutor has the background to understand them. To leverage this issue, we simplify text passages issued from computer science articles abstracts by a pair of experts. One annotator is a computer scientist who understands the text and simplifies passages. Then each pair of passages (simplified and not) is reread by a professional translator from the University of Western Brittany Translation Office[7] who is an English native speaker but not a specialist in computer science. Each passage is discussed and rewritten multiple times until it becomes clear for non computer scientists. The observation of the obtained simplification examples revealed opposite strategies in making text understandable. On the one hand, shortening passages by eliminating details and generalization

[5] https://www.aminer.org/citation.
[6] https://stellargraph.readthedocs.io/.
[7] https://www.univ-brest.fr/btu.

seem an efficient strategy. On the other hand, simplified sentences are longer and more concrete, e.g. the sentence from an article on exposing image tampering "The learning classifiers are applied for classification" was simplified as "The machine learning algorithms are applied to detect image manipulation". For a computer scientist, it is evident that the detection problem is a special case of a binary classification task, but in order to make this sentence understandable for a non computer scientist, the abstract term "classification" should be replaced with a concrete use-case "to detect image manipulation". Thus, on the one hand our methodology of passage simplification ensures data quality. On the other hand, it provides interesting insights to simplification strategies. 57 manually simplified passages were provided to participants for training.

We manually searched for difficult terms and ranked them from 1 to 10 according to their complexity. 1 corresponds to the terms that very difficult and unknown to the general public. Lower ranks shows that the term might be explained if there is a room. Notice, that the final ranking can be obtained by binary comparison of each pair of candidate terms.

We continue to simplify passages and search for difficult terms.

3.2 Queries

For this edition 13 queries are a selection of recent n press titles from *The Guardian* enriched with keywords manually extracted from the content of the article. It has been checked that each keyword allows to extract at least 5 relevant abstracts. The use of these keywords is optional.

Input format for all tasks:

- Topics in the MD format (see Fig. 1);
- Full text articles from The Guardian (link, folder query_related_content with full texts in the MD format);
- ElasticSearch index on the data server:[8];
- DBLP full dump in the JSON.GZ format;
- DBLP abstracts extracted for each topic in the following MD format (doc_id, year, abstract) (see Fig. 2).

[8] https://guacamole.univ-avignon.fr/nextcloud/index.php/apps/files/?dir=/simpleText/.

Query 1: Digital assistants like Siri and Alexa entrench gender biases, says UN

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

Topic 1.1: Digital assistant

https://inex:qatc2011@guacamole.univ-avignon.fr/dblp1/_search?q="Digital assistant"&size=1000

Topic 1.2: Biases

https://inex:qatc2011@guacamole.univ-avignon.fr/dblp1/_search?q=biases&size=1000

Fig. 1. Query example

1564531496	2002	In this short paper we describe the architectural cc
2988211052	2002	In this short paper we describe the architectural cc
3006661050	2003	Modern Personal Digital Assistant (PDA) architecture
1970213811	2006	This demonstration presents a new interaction techni
2797641221	2018	Digital assistants are emerging to become more preva
2158159346	2004	Abstract Mobile devices are significantly changing
2463945949	2016	DIANE is a digital assistant system that aims to fas

Fig. 2. DBLP abstract examples

4 Pilot Tasks

In 2021, SimpleText was run as a CLEF workshop. The goal was to create a community interested in generating a simplified summary of scientific documents and to define tasks and evaluation setup.

We proposed three pilot tasks to help to better understand the challenges as well as discuss these challenges and the way to evaluate solutions. Details on the tasks, guideline and call for contributions can be found at the SimpleText website[9], in this paper we just briefly introduce the planned pilot tasks. Note that the pilot tasks are means to help the discussions and to develop a research community around text simplification. Contributions are not exclusively rely on the pilot tasks.

43 teams were registered for the SimpleText workshop with 23 participants subscribed on our Google group and 24 followers on Twitter. Although data was downloaded from the server by several participants, they did not submit their runs on our pilot tasks due to the lack of time. We continue to enrich

[9] https://simpletext-madics.github.io/2021/clef/en/.

data prepared for the pilot tasks for the SimpleText@CLEF-2021 workshop to prepare an evaluation lab in 2022. As we did not perform evaluation this year, we present only potential evaluation metrics that can be used in the 2022 edition of SimpleText.

4.1 Task 1: Selecting Passages to Include in a Simplified Summary - Content Simplification

Given an article from a major international newspaper general audience, this pilot task aims at retrieving from a large scientific bibliographic database with abstracts, all passages that would be relevant to illustrate this article. Extracted passages should be adequate to be inserted as plain citations in the original paper.

Sentence pooling and automatic metrics can be used to evaluate these results. The relevance of the source document can be evaluated as well as potential unresolved anaphora issues.

Output: A maximum of 1000 passages to be included in a simplified summary in a TSV (Tab-Separated Values) file with the following fields:

- *run_id*: Run ID starting with *team_id_*;
- *manual*: Whether the run is manual 0,1;
- *topic_id*: Topic ID;
- *doc_id*: Source document ID;
- *passage*: Text of the selected passage;
- *rank*: Passage rank.

An output example is given in Table 1.

4.2 Task 2: Searching for Background Knowledge

The goal of this pilot task is to decide which terms (up to 10) require explanation and contextualization to help a reader to understand a complex scientific text - for example, with regard to a query, terms that need to be contextualized (with a definition, example and/or use-case). Terms should be ranked from 1 to 10 according to their complexity. *1* corresponds to the most difficult term, while lower ranks show that the term might be explained if there is a room.

Output: List of terms to be contextualized in a tabulated file TSV with the following fields:

- *run_id*: Run ID starting with *team_id_*;
- *manual*: Whether the run is manual 0,1;
- *topic_id*: Topic ID;
- *passage_text*: Passage text;
- *term*: Term or other phrase to be explained;
- *rank*: Importance of the explanation for a given term.

An output example for task 2 is given in Table 2.

Table 1. Task 1 output example

run_id	Manual	topic_id	doc_id	Passage	Rank
ST_1	1	1	3000234933	People are becoming increasingly comfortable using Digital Assistants (DAs) to interact with services or connected objects	1
ST_1	1	1	3003409254	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.	2
ST_1	1	1	3003409254	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	3

Term pooling and automatic metrics (NDCG, ...) will be used to evaluate these results in the future edition.

4.3 Task 3: Scientific Text Simplification

The goal of this pilot task is to provide a simplified version of text passages. Participants are provided with queries and abstracts of scientific papers. The abstracts can be split into sentences as in the example. The simplified passages will be evaluated manually with eventual use of aggregating metrics in the future edition.

Output: Simplified passages in a TSV tabulated file with the following fields:

– *run_id*: Run ID starting with *team_id_*;
– *manual*: Whether the run is manual 0,1;
– *topic_id*: Topic ID;
– *doc_id*: Source document ID;
– *source_passage*: Source passage text;
– *simplified_passage*: Text of the simplified passage.

An output example for task 3 is given in Table 3.

Table 2. Task 2 output example

run_id	Manual	topic_id	passage_text	Term	Rank
ST_1	1	1	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	Machine learning	1
ST_1	1	1	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	Societal biases	2
ST_1	1	1	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	ML	3

Table 3. Task 3 output example

run_id	Manual	topic_id	doc_id	source_passage	simplified_passage
ST_1	1	1	3003409254	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	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 dataset they use

5 Conclusion and Future Work

The paper introduced the CLEF 2021 SimpleText track, consisting of a workshop and pilot tasks on text simplification for scientific information access. Although 43 teams were registered for the SimpleText workshop and the data was downloaded from the server by several participants, they did not submit their runs on our pilot tasks due to the lack of time and therefore we did not perform evaluation this year. We continue to enrich data prepared for the tasks for the next edition of SimpleText.

The created collection of simplified texts makes it possible to apply overlap metrics like ROUGE to text simplification. However, we will work on a new evaluation metric that can take into account unresolved anaphora [4] and information types.

In future, we will perform deeper analysis of queries collected from different sources. We will reconsider source data: research papers/preprints and their abstracts (e.g. from HAL[10], arXiv[11], or ISTEX[12] platforms using unpaywall API[13] to search for open access versions), Wikipedia/SimpleWikipedia articles, science journalism articles (e.g. ScienceX[14] instead of The Guardian, as it can be freely shared for research purposes), forums like ELI5[15]. We will propose an evaluation lab at CLEF (instead of a workshop). The objective of the Task 1 will be to decide automatically which passages of the scientific articles/abstracts should be included in extractive summaries in order to get a simplified summary of the initial texts taking into account that the information in a summary designed for an expert should be different from that aimed at a general audience. For the pilot task 2, participants will be asked to provide context for difficult terms.

We will prepare datasets in French and enrich datasets in English. We will also propose baselines for all three tasks.

References

1. Aharoni, R., Goldberg, Y.: Split and rephrase: better evaluation and a stronger baseline. arXiv:1805.01035 [cs], May 2018. http://arxiv.org/abs/1805.01035
2. Anand Deshmukh, A., Sethi, U.: IR-BERT: leveraging bert for semantic search in background linking for news articles. arXiv e-prints **2007**. arXiv:2007.12603, July 2020. http://adsabs.harvard.edu/abs/2020arXiv200712603A
3. Augenstein, I., Das, M., Riedel, S., Vikraman, L., McCallum, A.: Semeval 2017 task 10: scienceie-extracting keyphrases and relations from scientific publications. arXiv preprint arXiv:1704.02853 (2017)

[10] https://hal.archives-ouvertes.fr/.
[11] https://arxiv.org/.
[12] https://istex.fr/.
[13] https://unpaywall.org/products/api.
[14] https://sciencex.com/.
[15] https://www.reddit.com/r/explainlikeimfive/.

4. Bellot, P., Moriceau, V., Mothe, J., SanJuan, E., Tannier, X.: INEX tweet contextualization task: evaluation, results and lesson learned. Inf. Process. Manage. **52**(5), 801–819 (2016). https://doi.org/10.1016/j.ipm.2016.03.002
5. Biran, O., Brody, S., Elhadad, N.: Putting it simply: a context-aware approach to lexical simplification. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 496–501. Association for Computational Linguistics, Portland, June 2011. https://www.aclweb.org/anthology/P11-2087
6. Botha, J.A., Faruqui, M., Alex, J., Baldridge, J., Das, D.: Learning to split and rephrase from Wikipedia edit history. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 732–737. Association for Computational Linguistics, Brussels, October 2018. https://doi.org/10.18653/v1/D18-1080. https://www.aclweb.org/anthology/D18-1080
7. Cardon, R., Grabar, N.: Détection automatique de phrases paralléles dans un corpus biomédical comparable technique/simplifié. In: TALN 2019, Toulouse, France, July 2019. https://hal.archives-ouvertes.fr/hal-02430446
8. Cardon, R., Grabar, N.: French biomedical text simplification: when small and precise helps. In: Proceedings of the 28th International Conference on Computational Linguistics, pp. 710–716. International Committee on Computational Linguistics, Barcelona, December 2020. https://doi.org/10.18653/v1/2020.coling-main.62. https://www.aclweb.org/anthology/2020.coling-main.62
9. Chen, P., Rochford, J., Kennedy, D.N., Djamasbi, S., Fay, P., Scott, W.: Automatic text simplification for people with intellectual disabilities. In: Artificial Intelligence Science and Technology, pp. 725–731. World Scientific, November 2016. https://doi.org/10.1142/9789813206823_0091. https://www.worldscientific.com/doi/abs/10.1142/97898132068230091
10. Collins-Thompson, K., Callan, J.: A language modeling approach to predicting reading difficulty. In: Proceedings of HLT/NAACL, vol. 4 (2004)
11. Coster, W., Kauchak, D.: Simple English Wikipedia: a new text simplification task. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 665–669 (2011)
12. Cram, D., Daille, B.: Terminology extraction with term variant detection. In: Proceedings of ACL-2016 System Demonstrations, pp. 13–18. Association for Computational Linguistics, Berlin, August 2016. https://doi.org/10.18653/v1/P16-4003. https://www.aclweb.org/anthology/P16-4003
13. Dernoncourt, F., Lee, J.Y.: PubMed 200k RCT: a dataset for sequential sentence classification in medical abstracts. In: Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers), pp. 308–313. Asian Federation of Natural Language Processing, Taipei, November 2017. https://www.aclweb.org/anthology/I17-2052
14. Dong, Y., Li, Z., Rezagholizadeh, M., Cheung, J.C.K.: EditNTS: an neural programmer-interpreter model for sentence simplification through explicit editing. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 3393–3402. Association for Computational Linguistics, Florence, July 2019. https://doi.org/10.18653/v1/P19-1331. https://www.aclweb.org/anthology/P19-1331
15. Ermakova, L., et al..: Text simplification for scientific information access: CLEF 2021 simpletext workshop. In: Proceedings of Advances in Information Retrieval - 43nd European Conference on IR Research, ECIR 2021, Lucca, Italy, 28 March–1 April 2021. Lucca, Italy (2021)

16. Ermakova, L., Bordignon, F., Turenne, N., Noel, M.: Is the abstract a mere teaser? Evaluating generosity of article abstracts in the environmental sciences. Front. Res. Metrics Anal. **3** (2018). https://doi.org/10.3389/frma.2018.00016. https://www.frontiersin.org/articles/10.3389/frma.2018.00016/full
17. Ermakova, L., Cossu, J.V., Mothe, J.: A survey on evaluation of summarization methods. Inf. Process. Manage. **56**(5), 1794–1814 (2019). https://doi.org/10.1016/j.ipm.2019.04.001. http://www.sciencedirect.com/science/article/pii/S0306457318306241
18. Ermakova, L., Goeuriot, L., Mothe, J., Mulhem, P., Nie, J.Y., SanJuan, E.: CLEF 2017 microblog cultural contextualization lab overview. In: Proceedings of Experimental IR Meets Multilinguality, Multimodality, and Interaction - 8th International Conference of the CLEF Association, CLEF 2017, Dublin, Ireland, 11–14 September 2017,pp. 304–314 (2017). https://doi.org/10.1007/978-3-319-65813-1_27
19. Fang, F., Stevens, M.: Sentence simplification with transformer-XL and paraphrase rules, p. 10 (2019)
20. Fecher, B., Friesike, S.: Open science: one term, five schools of thought. In: Bartling, S., Friesike, S. (eds.) Opening Science, pp. 17–47. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-00026-8_2
21. Flesch, R.: A new readability yardstick. J. Appl. Psychol. **32**(3), 221–233 (1948)
22. Fontelo, P., Gavino, A., Sarmiento, R.F.: Comparing data accuracy between structured abstracts and full-text journal articles: implications in their use for informing clinical decisions. Evidence-Based Med. **18**(6), 207–11 (2013). https://doi.org/10.1136/eb-2013-101272. http://www.researchgate.net/publication/240308203_Comparing_data_accuracy_between_structured_abstracts_and_full-text_journal_articles_implications_in_their_use_for_informing_clinical_decisions
23. Glavaš, G., Štajner, S.: Simplifying lexical simplification: do we need simplified corpora? In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). pp. 63–68. Association for Computational Linguistics, Beijing, July 2015. https://doi.org/10.3115/v1/P15-2011. https://www.aclweb.org/anthology/P15-2011
24. Grabar, N., Cardon, R.: CLEAR-simple corpus for medical French, November 2018. https://halshs.archives-ouvertes.fr/halshs-01968355
25. Hoffart, J., et al.: Robust disambiguation of named entities in text. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pp. 782–792 (2011)
26. Jarreau, P.B., Porter, L.: Science in the social media age: profiles of science blog readers. J. Mass Commun. Q. **95**(1), 142–168 (2018). https://doi.org/10.1177/1077699016685558
27. Jiang, C., Maddela, M., Lan, W., Zhong, Y., Xu, W.: Neural CRF model for sentence alignment in text simplification. arXiv:2005.02324 [cs], June 2020. http://arxiv.org/abs/2005.02324
28. Jin, D., Szolovits, P.: Hierarchical neural networks for sequential sentence classification in medical scientific abstracts. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 3100–3109. Association for Computational Linguistics, Brussels, October 2018. https://doi.org/10.18653/v1/D18-1349. https://www.aclweb.org/anthology/D18-1349

29. Kauchak, D.: Improving text simplification language modeling using unsimplified text data. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1537–1546. Association for Computational Linguistics, Sofia, August 2013. https://www.aclweb.org/anthology/P13-1151
30. Leroy, G., Endicott, J.E., Kauchak, D., Mouradi, O., Just, M.: User evaluation of the effects of a text simplification algorithm using term familiarity on perception, understanding, learning, and information retention. J. Med. Internet Res. **15**(7), e144 (2013)
31. Lin, C.Y.: ROUGE: a package for automatic evaluation of summaries. In: Text Summarization Branches Out: Proceedings of the ACL–04 Workshop, pp. 74–81 (2004)
32. Liu, Y., Lapata, M.: Text summarization with pretrained encoders. arXiv:1908.08345 [cs], September 2019. http://arxiv.org/abs/1908.08345
33. Maddela, M., Alva-Manchego, F., Xu, W.: Controllable text simplification with explicit paraphrasing. arXiv:2010.11004 [cs], April 2021.http://arxiv.org/abs/2010.11004
34. Maddela, M., Xu, W.: A word-complexity lexicon and a neural readability ranking model for lexical simplification. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 3749–3760. Association for Computational Linguistics, Brussels (2018). https://doi.org/10.18653/v1/D18-1410. https://www.aclweb.org/anthology/D18-1410
35. Martin, L., et al.: CamemBERT: a tasty French language model. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 7203–7219. Association for Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.acl-main.645. https://www.aclweb.org/anthology/2020.acl-main.645
36. Maruyama, T., Yamamoto, K.: Extremely low resource text simplification with pre-trained transformer language model. In: International Conference on Asian Language Processing p. 6 (2019)
37. McCarthy, P.M., Guess, R.H., McNamara, D.S.: The components of paraphrase evaluations. Behav. Res. Methods **41**(3), 682–690 (2009). https://doi.org/10.3758/BRM.41.3.682. https://doi.org/10.3758/BRM.41.3.682
38. Michalsky, T.: When to scaffold motivational self-regulation strategies for high school students' science text comprehension. Front. Psychol. **12** (2021). https://doi.org/10.3389/fpsyg.2021.658027. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.658027/full
39. Molek-Kozakowska, K.: Communicating environmental science beyond academia: stylistic patterns of newsworthiness in popular science journalism. Discour. Commun. **11**(1), 69–88 (2017). https://doi.org/10.1177/1750481316683294
40. Narayan, S., Gardent, C., Cohen, S.B., Shimorina, A.: Split and rephrase. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 606–616. Association for Computational Linguistics, Copenhagen, September 2017. https://doi.org/10.18653/v1/D17-1064. https://www.aclweb.org/anthology/D17-1064
41. Nenkova, A., Passonneau, R., McKeown, K.: The pyramid method: incorporating human content selection variation in summarization evaluation. ACM Trans. Speech Lang. Process. **4**(2) (2007). https://doi.org/10.1145/1233912.1233913
42. Owczarzak, K., Dang, H.T.: Overview of the TAC 2011 summarization track: guided task and AESOP task. In: Proceedings of the Text Analysis Conference (TAC 2011), Gaithersburg, Maryland, USA, November 2011

43. O'Reilly, T., Wang, Z., Sabatini, J.: How much knowledge is too little? When a lack of knowledge becomes a barrier to comprehension. Psychol. Sci. (2019). https://doi.org/10.1177/0956797619862276. https://journals.sagepub.com/doi/10.1177/0956797619862276
44. Paetzold, G., Specia, L.: Lexical simplification with neural ranking. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, pp. 34–40. Association for Computational Linguistics, Valencia, April 2017. https://www.aclweb.org/anthology/E17-2006
45. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: BLEU: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 311–318. Association for Computational Linguistics (2002)
46. Rao, S., Tetreault, J.: Dear sir or madam, may i introduce the GYAFC dataset: corpus, benchmarks and metrics for formality style transfer. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pp. 129–140 (2018)
47. Sadoski, M.: Reading comprehension is embodied: theoretical and practical considerations. Educ. Psychol. Rev. **30**(2), 331–349 (2018). https://doi.org/10.1007/s10648-017-9412-8
48. Si, L., Callan, J.: A statistical model for scientific readability. In: Proceedings of the Tenth International Conference on Information and Knowledge Management, CIKM 2001, pp. 574–576. ACM, New York (2001). https://doi.org/10.1145/502585.502695
49. Specia, L., Jauhar, S.K., Mihalcea, R.: SemEval-2012 task 1: english lexical simplification. In: *SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012), pp. 347–355. Association for Computational Linguistics, Montréal (2012). https://www.aclweb.org/anthology/S12-1046
50. Sulem, E., Abend, O., Rappoport, A.: Semantic structural evaluation for text simplification. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pp. 685–696 (2018)
51. Wang, T., Chen, P., Rochford, J., Qiang, J.: Text simplification using neural machine translation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 30, no. 1, March 2016. https://ojs.aaai.org/index.php/AAAI/article/view/9933
52. Wang, W., Li, P., Zheng, H.T.: Consistency and coherency enhanced story generation. arXiv:2010.08822 [cs], October 2020. http://arxiv.org/abs/2010.08822
53. Woodsend, K., Lapata, M.: Learning to simplify sentences with quasi-synchronous grammar and integer programming. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pp. 409–420. Association for Computational Linguistics, Edinburgh, July 2011. https://www.aclweb.org/anthology/D11-1038
54. Wubben, S., van den Bosch, A., Krahmer, E.: Sentence simplification by monolingual machine translation. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1015–1024 (2012)

55. Xu, W., Callison-Burch, C., Napoles, C.: Problems in current text simplification research: new data can help. Trans. Assoc. Comput. Linguist. **3**, 283–297 (2015). https://doi.org/10.1162/tacl_a_00139. https://www.mitpressjournals.org/doi/abs/10.1162/tacla00139
56. Xu, W., Napoles, C., Pavlick, E., Chen, Q., Callison-Burch, C.: Optimizing statistical machine translation for text simplification. Trans. Assoc. Comput. Linguist. **4**, 401–415 (2016)
57. Yatskar, M., Pang, B., Danescu-Niculescu-Mizil, C., Lee, L.: For the sake of simplicity: unsupervised extraction of lexical simplifications from Wikipedia. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 365–368. Association for Computational Linguistics, Los Angeles, June 2010. https://www.aclweb.org/anthology/N10-1056
58. Zhao, S., Meng, R., He, D., Saptono, A., Parmanto, B.: Integrating transformer and paraphrase rules for sentence simplification. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 3164–3173. Association for Computational Linguistics, Brussels, October 2018. https://doi.org/10.18653/v1/D18-1355. https://www.aclweb.org/anthology/D18-1355
59. Zhong, Y., Jiang, C., Xu, W., Li, J.J.: Discourse level factors for sentence deletion in text simplification. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 05, pp. 9709–9716, April 2020. https://doi.org/10.1609/aaai.v34i05.6520. https://ojs.aaai.org/index.php/AAAI/article/view/6520
60. Zhu, Z., Bernhard, D., Gurevych, I.: A monolingual tree-based translation model for sentence simplification. In: Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), pp. 1353–1361. Coling 2010 Organizing Committee, Beijing, August 2010. https://www.aclweb.org/anthology/C10-1152
61. Štajner, S., Nisioi, S.: A detailed evaluation of neural sequence-to-sequence models for in-domain and cross-domain text simplification. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA), Miyazaki, Japan, May 2018. https://www.aclweb.org/anthology/L18-1479