Towards Computer-aided Analysis of Readability and Comprehensibility of Patient Information in the Context of Clinical Research Projects

Ingo Glaser
Technical University of Munich
Garching bei München, Germany
ingo.glaser@tum.de

Georg Bonczek
Technical University of Munich
Garching bei München, Germany
georg.bonczek@tum.de

Jörg Landthaler
Technical University of Munich
Garching bei München, Germany
joerg.landthaler@tum.de

Florian Matthes
Technical University of Munich
Garching bei München, Germany
florian.matthes@tum.de

1 INTRODUCTION

Clinical trials aim to examine the safety and efficiency of novel procedures such as certain medication, specific treatments, or medical devices. They constitute a compulsory requirement for the marketing authorization of these trial objects. As such, they are a vital part of a modern and innovative health care system. However, clinical trials (CT) also involve risks for the participant's health.

To give the patient the right to self-determination, the Declaration of Helsinki [1] — one of the most influential set of ethical principles regarding human experimentation — bases its principles on the concept of *informed consent*. This means that participants are required to voluntarily give consent to partake in the study after they have been adequately informed about all relevant aspects potentially influencing their treatment. The ideas of the Declaration of Helsinki have influenced many national or regional legislation. In practice, potential participants will usually be given a written document containing all relevant information on the basis of which their informed consent should be ensured. This document is called patient information (PI). In Germany, before a clinical trial involving patients can commence, the approval of the responsible Ethics Commission (EC) is necessary. This also includes a review of the PI. Even though the participants' health that is at stake makes the EC apply rigorous scrutiny during that task, the distributed information material is often still hard to understand. Several reasons lead to this disconnect. Firstly, the analysis of the often voluminous PI is a time consuming process, yet the PI constitutes only a fraction of the complete application that has to be approved and it also has to be processed in limited time. Additionally, the lack of readability of the PI is not always considered a sufficient reason for the EC to disapprove of the clinical trial and consequently deny the application. If the patient cannot understand the implications of their consent, it is questionable if they can even give their *informed* consent. Still including patients in the trial would constitute a violation of basic ethical principles. Therefore, an automatic validation of many of the requirements such as the readability of the document, would greatly improve the efficiency of the EC's review process, enabling the patients to make better informed and self-determined decisions. Apart from the patient and the EC, the applicant would also profit from such a system as they do not have to wait for feedback, but are instantly engaged by intelligent tools that aid their writing and application process. Hence, in this paper we propose a system that automatically checks the readability and comprehensibility of PI to encourage the creator of the documents to write more understandable information material by giving explicit suggestions on how to improve the quality.

ICAIL '19, June 17–21, 2019, Montreal, QC, Canada

ACM ISBN 978-1-4503-6754-7/19/06.
https://doi.org/10.1145/3322640.3326704

2 REQUIREMENTS ON PATIENT INFORMATION

Two main sources — a checklist of 60 (sometimes optional) requirements [5] and 100 additional demands of an ethics commission regarding the quality of patient information — were used to identify common requirements concerning readability and comprehensibility. In 70 of the 100 additional demands, the PI is mentioned, and in 35 of these 70 some concerns regarding readability were raised. Most of them pointing out the lack of comprehensibility to a layperson. This includes basic problems such as sentence or document length, long enumerations or complex tables. The more specific remarks revolve around the usage of technical or foreign words. While some words (e.g. specific legal terms) should be completely omitted, others should only occur in a preferred version (e.g. 'Ethikkommission' instead of 'Ethikausschuss'). According to three domain experts, medical terms should be avoided when possible, otherwise an explanation needs to be added. Often, the explanations are wrong, not adequate or certain terms are used before they have been introduced to the patient. In some cases medical conditions or procedures require further information, for example when a patient is diagnosed with a specific disease which requires reporting to authorities, they needs to know about this obligation. Therefore, this information needs to be included in the PI as well.

Firstly, the presented system shall be able to identify overly long sentences and measure the readability in some generic way. While

being trivial to implement, these metrics might be able to give a first impression of the readability of the document.

In a second part, the system shall be able to identify medical technical terms and detect whether they have been explained. However, only the terms that might not be understandable to a general audience should be detected. If the detection of terms and explanations is implemented, more advanced rules can be automatically forced on the documents (e.g. detecting whether a term has been used before its definition).

3 PROPOSED SYSTEM

While we suggest different solutions for the problems described in this paper, all of them share a common Apache UIMA pipeline to connect different NLP components, which gradually add more information by creating annotations on the input text. The first component carries out different generic NLP preprocessing steps. Afterwards, a readability metrics component is used to assess the readability of the present PI, by means of different readability measures (e.g. Flesch-Reading-Ease [3], gSMOG [6] or Wiener Sachtextformel [2]).

The main task of the proposed system, detecting technical terms, is twofold. Technical terms are acceptable to some extend as long as they are explained. This means a complete system would have to both identify technical terms and recognize whether they have been explained. In a first step, we only care about *detecting* such entities, not their disambiguation or determining whether the term has been explained. Therefore, named entity recognition (NER) can be used. Due to the lack of freely available biomedical NER implementations for German text, our system proposes a recognizer based on medical dictionaries. We suggest multiple German dictionaries as knowledge base, namely: (1) ICD10, (2) UMDNS, (3) ICPC, (4) ICDO3, (5) OPS, (6) LOINC, (7) MEddra, and (8) Mesh. For the actual NER, different approaches could be used, such as longest prefix matching, locality sensitive sashing [9] (LSH), SimString [7], or MetaMapLite [8].

Not all terms found in the used dictionaries are unfamiliar to readers (e.g. 'face' or 'grandparent'), and also, not all unfamiliar terms need to be of biomedical nature (e.g. 'personal data'). To address this issue, we add another processing stage after the entity recognition. Each entity found is checked whether it is a commonly used term, if this is the case, it is removed from the list of candidates. To determine the frequency of a word, the 'News' corpus of the Leipzig corpora collection [4] will be used. This corpus of 1 million words crawled from encodes the absolute number of occurrences of each word. Based on this we can compute the frequency class as $C_w = \log_2 \frac{F_h}{F_w}$ where F_h is the frequency of the most common word and F_w is the frequency of the current word. Using this C, we can reason about the reader's familiarity with a word. By dynamically choosing a threshold T, such that a word w is included in the final candidate set only if $C_w \geq T$, the amount of terms annotated as too complex can be controlled, matching the expected vocabulary of the patient targeted by the PI. In the biomedical domain, companies often use medical terms or variations thereof, in their company names. To exclude these terms, every technical term found which is contained in NER-tags of type *ORGANIZATION*, should be removed from the list of candidates.

To detect whether a technical term has been explained, we suggest multiple steps to first detect all candidates for explanations and then filtering them afterwards. In a first step, possible explanations of each term could searched by a set of rules, including regular expressions. For example, explanations are often given in parenthesis after the term. The filtering process then in turn consists of multiple conditions. As first condition, every explanation that is in turn a single technical term or an abbreviation of the technical term, can be removed from the set of candidates. A second condition ensures that only a single explanation is linked to a term. If multiple explanations are detected (either because the term is actually explained multiple times or a rule triggered a false positive), the first one found is chosen. In a last step, circular term-explanation references are resolved. One type of explanation is of the form 'explanation (term)', which can also be in the form 'term (explanation)'. When both the explanation and the term are detected as technical terms, a circular reference occurs, as the first explanation acts as term in the second form and the first term acts as the explanation in the second form. Such a state is resolved by using the technical term with the lowest frequency class as the explanation and the other one as the targeted term. Every occurrence of the technical term is then linked to that explanation. From this point on, a term is considered *explained* if an explanation is linked to it.

4 CONCLUSION

In this work, we introduced the problem of unreadable patient information and their implications on informed consent in clinical trials. We presented some common mistakes leading to unreadable and incomprehensible documents, derived from actual patient information. A system that automatically check for these types of errors was proposed. This includes general readability metrics to quickly assess the general quality of the document, while named entity recognizers can be used to extract overly technical terms and their explanations and other information in relation with these terms.

5 ACKNOWLEDGEMENT

This research was sponsored in part by the German Federal Ministry of Education and Research (BMBF) (project "Software Campus (TU München)", grant no. 01IS17049).

REFERENCES

[1] World Medical Association. 2013. World medical association declaration of helsinki: Ethical principles for medical research involving human subjects. *The Journal of the American Medical Association* 310, 20 (2013), 2191–2194. https://doi.org/10.1001/jama.2013.281053

[2] Richard Bamberger and Erich Vanecek. 1984. *Lesen - Verstehen - Lernen - Schreiben. Die Schwierigkeitsstufen von Texten in deutscher Sprache.* Jugend u. Volk Sauerlaender, Vienna, Austria.

[3] Rudolph Flesch. 1948. A new readability yardstick. *Journal of applied psychology* 32, 3 (1948), 221.

[4] Dirk Goldhahn, Thomas Eckart, and Uwe Quasthoff. 2012. Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages. In *LREC*, Vol. 29. 31–43.

[5] Urs Harnischmacher, Peter Ihle, Bettina Berger, Jürgen Goebel, and Jürgen Scheller. 2006. *Checkliste und Leitfaden zur Patienteneinwilligung.* Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin, Germany.

[6] G Harry Mc Laughlin. 1969. SMOG grading-a new readability formula. *Journal of reading* 12, 8 (1969), 639–646.

[7] Naoaki Okazaki and Jun'ichi Tsujii. 2010. Simple and Efficient Algorithm for Approximate Dictionary Matching. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*. Association for Computational Linguistics, Beijing, China, 851–859. http://www.aclweb.org/anthology/C10-1096

[8] Dina Demner-Fushman Willie J. Rogers and Alan R. Aronson. 2017. MetaMap Lite: an evaluation of a new Java implementation of MetaMap. *Journal of the American Medical Informatics Association* 24, 4 (2017), 841–844.

[9] Jingdong Wang, Heng Tao Shen, Jingkuan Song, and Jianqiu Ji. 2014. Hashing for Similarity Search: A Survey. *arXiv e-prints*, Article arXiv:1408.2927 (Aug. 2014), arXiv:1408.2927 pages. arXiv:cs.DS/1408.2927