International Journal of Data Warehousing and Mining
Volume 16 • Issue 1 • January-March 2020

Patient Oriented Readability Assessment for Heart Disease Healthcare Documents

Hui-Huang Hsu, Tamkang University, New Taipei City, Taiwan
Yu-Sheng Chen, National Taiwan Ocean University, Keelung, Taiwan
Chuan-Jie Lin, National Taiwan Ocean University, Keelung, Taiwan
Tun-Wen Pai, National Taipei University of Technology, Taipei, Taiwan

ABSTRACT

Personal health literacy is an important indicator for a national health status. Providing citizens with sufficient medical knowledge can help them understand their own health conditions. To achieve this goal, an integrated system is developed for evaluating the readability of healthcare documents by taking heart disease as a specific topic. The mechanism can be extended to other target diseases and languages by changing the corresponding word databank. The assessment system for examining document readability is based on patient-oriented aspects rather than professional aspects. Commonly used terms and professional medical terms extracted from a query document were utilized as fundamental elements for readability analysis, and the derived features included term frequency of professional medical terms, proportion of professional medical terms, and diversity indicator of medical terms. A five-fold cross validation is applied to measure the robustness of the proposed approach. The experimental results achieved a recall rate of 0.93, a precision rate of 0.97, and an accuracy rate of 0.95.

KEYWORDS

Cardiovascular, Diversity Indicator, Medical Care, Readability, Word Frequency

1. INTRODUCTION

With proper analytical mechanisms, digitalized clinical/medical documents can be very helpful in disease diagnosis and treatments (Hussain & Lee, 2015; Barbantan et al., 2016). Due to the vigorous development of internet technology, an abundant variety of online healthcare-related documents can also be accessed easily. They are particularly important resources for patients and their families to learn medical knowledge for home self-care. However, most healthcare documents do not go through objective testing to determine whether the document is suitable for self-learning by patients. Accurate and understandable medical knowledge is especially important for patients after surgery, so they can be educated about their recovery status, learn self-care, and respond to emergency events. With improvement in personal health literacy, patients can cooperate with professional medical treatments, increase the selectivity of possible medical therapies, extend the duration of re-hospitalization, and reduce medical expenses.

Readability of a document is to measure the ease with which a reader can understand the contents of the document. There have been many research works in the NLP community focusing on document readability assessment. Many studies proposed to evaluate readability degrees by different kinds of features, such as vocabulary, ngrams, sentence length, entity-density, syntax, POS, semantics, and

DOI: 10.4018/IJDWM.2020010104

Copyright © 2020, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

co-reference features (Heilman et al., 2007; Feng et al., 2010; Nenkova et al., 2010; Clercq & Hoste, 2016). All the features were primarily designed for English texts. Traditionally, readability was measured by specifically designed formulas (Dubay, 2004; Wang et al., 2012). Recent works also applied different machine learning techniques to build readability predictors (Schwarm & Ostendorf, 2005; Kate et al., 2010; Mukherjee et al., 2018), and Lasecki et al. also discussed the possibility of measuring by evaluating crowd features (Lasecki et al., 2015).

The readability of different kinds of medical documents has received much attention in the last decade (Miller et al., 2007; Leroy & Endicott, 2011; Atcherson, 2013; Williamson & Martinez, 2013). Kauchak et al. found that specificity features (calculated using word-level depth in MeSH) and ambiguity features (calculated using the number of UMLS Metathesaurus concepts associated with a word) were the strongest predictors for English documents (Kauchak et al., 2014). There are also different approaches for measuring the readability of general articles (Flesch, 1948; Fry, 2006). In addition, several readability assessment tools exist that are specifically for online medical education documents (Kher et al., 2017; Colaco et al., 2013).

The research mainly focused on documents related to dysfunctional heart failure. Researchers found that only 7.1% of the assessable issue-specific online medical documents passed the readability test, indicating that other online medical documents may not be suitable for the general public. Therefore, those responsible for creating medical education documents should determine whether the document readability is at an ideal level for public understanding.

This study aimed to develop an assessment tool for evaluating the readability of healthcare documents based on patient-oriented aspects, and the final goal was to provide an automatic evaluation system of document filters that can review all disease-specific medical education documents. However, due to the tremendous requirements for developing a large-scale, comprehensive system, this research illustrated the concept by focusing solely on medical education documents related to heart disease. The proposed automatic evaluation mechanism is expected to effectively distinguish the readability of healthcare documents regarding heart disease issues. Through the readability assessment and inspection, authors can rethink their writing goals and revise their previous written healthcare documents based from the perspective of patients.

2. MATERIAL AND METHOD

The automatic evaluation technology for document readability helped readers/writers assess whether the written article contents were suitable for the public. The heuristic evaluation approach used was observing whether a document contained common vocabulary of daily life and reduced the usage of professional terms. The system configuration of the automatic assessment processes for document readability can be divided into two main modules as shown in Figure 1. The first module was to determine whether the content of an article was related to the domain-specific theme, and concentrated on a single specific topic. Addressing a single topic in a medical education document can effectively improve the readability of the document. The second module was to calculate proposed features regarding article readability and to perform readability classification. Detailed descriptions of the two modules are provided in 2.1 and 2.2.

2.1 Module I: Domain-specific Theme Detection

The subject of the healthcare documents in this study focused on heart diseases. The main approach was to compare words used in a query document with the previously collected word databank related to heart disease. If the query document contained any key terms associated with heart disease, the module determined the document to be a domain-specific healthcare document and proceeded to the readability test.

The keyword databank is a collection of heart disease terminologies from seven professional medical institutions websites: Cincinnati Children's (2016, Nov. 20), Cleveland Clinic (2018, Nov.

8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/patient-oriented-readability-assessment-for-heart-disease-healthcare-documents/243414

Related Content

Cost Models for Selecting Materialized Views in Public Clouds
Romain Perriot, Jérémy Pfeifer, Laurent d'Orazio, Bruno Bachelet, Sandro Bimonteand Jérôme Darmont (2014). *International Journal of Data Warehousing and Mining (pp. 1-25).*
www.irma-international.org/article/cost-models-for-selecting-materialized-views-in-public-clouds/117156

Hierarchical Hybrid Neural Networks With Multi-Head Attention for Document Classification
Weihao Huang, Jiaojiao Chen, Qianhua Cai, Xuejie Liu, Yudong Zhangand Xiaohui Hu (2022). *International Journal of Data Warehousing and Mining (pp. 1-16).*
www.irma-international.org/article/hierarchical-hybrid-neural-networks-with-multi-head-attention-for-document-classification/303673

Machine Learning Based Admission Data Processing for Early Forecasting Students' Learning Outcomes
Nguyen Thi Kim Son, Nguyen Van Bien, Nguyen Huu Quynhand Chu Cam Tho (2022). *International Journal of Data Warehousing and Mining (pp. 1-15).*
www.irma-international.org/article/machine-learning-based-admission-data-processing-for-early-forecasting-students-learning-outcomes/313585

Discovering Surprising Instances of Simpson's Paradox in Hierarchical Multidimensional Data
Carem C. Fabrisand Alex A. Freitas (2006). *International Journal of Data Warehousing and Mining (pp. 27-49).*
www.irma-international.org/article/discovering-surprising-instances-simpson-paradox/1762

Determination of Unithood and Termhood for Term Recognition
Wilson Wong (2009). *Handbook of Research on Text and Web Mining Technologies (pp. 500-529).*
www.irma-international.org/chapter/determination-unithood-termhood-term-recognition/21743