



Symposium on Data Mining Applications, SDMA2016, 30 March 2016, Riyadh, Saudi Arabia

Readability of Arabic Medicine Information Leaflets: A Machine Learning Approach

Sihaam Alotaibi^a, Maha Alyahya^a, Hend Al-Khalifa^a, Sinaa Alageel^b and Nora Abanmy^b

^aCollege of Computer and Information Science, King Saud University, Riyadh ,Saudi Arabia

^bCollege of Pharmacy, King Saud University, Riyadh ,Saudi Arabia

Abstract

This paper presents a project that explores the possibility of assessing the readability level of Arabic medicine information leaflets using machine learning techniques. There are a number of popular readability formulas and tools that have been successfully used to assess the readability of health-related information in several languages. However, there is limited work on the readability assessment of health-related information, specifically medicine information leaflets in Arabic. We describe the design of a tool that uses machine learning to assess the readability of medicine information leaflets. We utilize a corpus comprising 1112 medicine information leaflets annotated with three difficulty levels. Based on a study of existing literature, we selected a number of features influencing text difficulty. The tool will help specialized organizations in medicine information leaflets production to produce the leaflets at appropriate level of reading for the majority of leaflets consumers.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the Organizing Committee of SDMA2016

Keywords: Text Readability; Medicine Information Leaflets; Machine Learning; Arabic Language; Arabic Natural Language Processing.

1. Introduction

Text readability is described as “the ease of understanding or comprehension because of the writing style”¹. Text readability assessment provides a measure of the text appropriateness to target readers. It provides a number of potential benefits in many fields such as education and healthcare. In education, it helps teachers match between texts and the reading abilities of students. In healthcare, it helps to provide health-related information such as medical instructions understood by the average patient.

* Corresponding author. Tel.: +966532221149 .

E-mail address: simalmish@gmail.com

There are a number of popular readability formulas and tools that have been successfully used to assess the readability of health-related information in several languages^{2, 3}. To the best of our knowledge, there have been a number of researches on automatic Arabic readability measurements of different types of documents such as curriculum textbooks and web articles^{4, 5, 6}, however, no existing research has explored the readability of health-related information, specifically medicine information leaflets in Arabic. Medicine Information leaflets or package inserts are considered an important source of health-related information. Medicine Information leaflets are leaflets including information about medical conditions, doses, side effects that packed with medicines to give the user more information about the medicine.

1.1. Problem definition

In Saudi Arabia, a survey of over 2000 community pharmacy customers found that the number of respondents who read package inserts (PIs), or at least ask someone else to read the PIs exceeds 85%⁷. This indicates that PIs sometimes become the main source of information for some patients for many reasons⁸. First, self-medication is common in Saudi Arabia⁹. Second, the purchase of prescription-only medications (POM) without community pharmacies prescription is common in Saudi Arabia¹⁰. This means that the patient assumes the bulk of responsibility for medication safety not only for over-the-counter (OTC) medications but also for medications such as antibiotics, nonsteroidal anti-inflammatories, and oral contraceptives⁸. Third, the information provided by physicians or pharmacists that describe dose and frequency, precautions, and adverse effects of their prescribed medications has been shown to be suboptimal¹¹.

Challenges in reading and understanding such written medicine information leaflets may represent one cause for the high rates of medication errors, such as poor adherence which attenuates optimum medicine benefit. To eliminate this challenge it is necessary to find ways to improve the readability of these leaflets and one such method is based on the assessment of text readability. The main goal of this project is to develop an Arabic text readability assessment tool for consumers of medicine information leaflets. Therefore, our project aims to answer the following question:

Given some medicine information leaflets that are written in Arabic, can a readability level of the text be measured automatically and accurately, and to what extent is it comparable to human judgment?

2. Literature review

Researchers have tried to use several approaches and features that help to achieve an accurate estimate of the readability level for health-related texts when developing readability assessment tools. Several studies have followed the traditional approach which is based on formulas^{12, 13}. The researchers found that readability formulas are not specifically tailored for medical content¹⁴. However, these formulas revealed to serve as an accepted initial approximation for predicting the consumer's comprehension of health topics¹⁵. Machine learning techniques such as Support Vector Machine, Decision Trees and Naïve Bayes were used in the development of a model that assesses the readability of web-based health information in a number of studies^{2, 16}.

In addition, the readability assessment research that uses computational approach is nearly new for the Arabic language. It has been applied on different types of documents such as curriculum textbooks and web articles^{4, 5, 24}. However, there is a lack of computational approaches for the purpose of readability assessment of health-related text in Arabic language.

3. Methodology

In this section, we will describe the tools used in our readability assessment tool, to help to generate the statistical language models and accomplish readability assessment task. Then we will describe the readability assessment tool development phases.

Download English Version:

<https://daneshyari.com/en/article/488583>

Download Persian Version:

<https://daneshyari.com/article/488583>

[Daneshyari.com](https://daneshyari.com)