Chapter 7.28
Ontology–Based Spelling Correction for Searching Medical Information

Jane Moon
Monash University, Australia

Frada Burstein
Monash University, Australia

ABSTRACT

There has been a paradigm shift in medical practice. More and more consumers are using the Internet as a source for medical information even before seeing a doctor. The well known fact is that medical terms are often hard to spell. Despite advances in technology, the Internet is still producing futile searches when the search terms are misspelled. Often consumers are frustrated with irrelevant information they retrieve as a result of misspelling. An ontology-based search is one way of assisting users in correcting their spelling errors when searching for medical information. This chapter reviews the types of spelling errors that adults make and identifies current technology available to overcome the problem.

INTRODUCTION

Medical terminology includes many words which nonprofessionals find difficult to spell. For novice computer users it can be extremely frustrating when, as a result of misspelling, they cannot find relevant information. The medical implications of errors arising as a result of misspelling a word is well documented in the literature (Lambert, 1997; Lilley, 1995). Reports show that there is a significant error rate observed in patient records—in particular, in discharge letters (Blaser et al., 2004). An intelligent system could provide the users with some suggestions in order to assist in using correct words when dealing with medical information. However; it seems we are still far from reaching this goal.

A word processing program such as Microsoft Word, auto corrects typos and provides advice on whether a word is correctly spelt through a

process of "verification and offers users to spell correctly via a process of 'correction'" (Berghel & Andreu, 1998). However, word correction is based on text spelling alone and can cause more problems at the correction stage. For example, Microsoft Word does not distinguish homophones such as *heroin* or *heroine* (Jones & Martin, 1997). The approach it takes in correcting spelling is not context based (Fallman, 2002) or ontology based (Khan, McLeod, & Hovy, 2004; Patel, Supekar, Lee, & Park, 2003; Tijerino & Reza, 2005), it does not recognize blank space which wastes a lot of space (Nix, 1981) and does not exhibit much intelligence. Popular search engines such as Yahoo and Google do provide some spelling correction by prompting a list of words based on keywords. In most cases search method is based on simple word searches and frequency distributions and these do not capture the meaning behind the words. Mauldin (1991) calls this phenomenon *keyword barrier.* Breaking through the keyword barrier will require a system that understands the texts they process.

There have been many attempts to produce spelling correction programs. They focus on various approaches including word concatenation (Berghel & Andreu, 1998), spelling correctors such as *Talisman* (Berghel, 1998), and correcting misspellings that contain a single error form (Pollock & Zamora, 1984). Some of these systems are domain specific and operate in chemistry, such as ChemSpell (Mitton, 1996; Zamora, 1978, 1980).

With the recent explosion of portals and ever expanding health information searching via medical/health portals, the question of the effectiveness of these health portals remains to be answered. Despite major advances and efforts to make portals user friendly, the research shows (Moon, 2005; Moon & Burstein, 2005) that most portals still lack essential functions for assisting users with information retrieval and hence losing potential users from existing portals. Even though the medical portals analyzed had the answers

to users' queries, the major problem lay in the portal's inability to handle misspelling, thereby lacking an essential provision for assisting users (Moon & Fisher, 2005).

Another issue is that some major search engines provide a list of words that are not, however, always relevant to the user's query. This is demonstrated in Figures 2-4. Recent advances in ontology-based search is suggested to be overcoming the shortcomings of keyword search (Khan et al., 2004). At least it eliminates lists of words that are not domain specific. It is highly domain specific and context based, thereby eliminating unnecessary retrieval of information. It is an efficient, time-saving method.

In this research we explored the basic patterns of adult spelling errors and researched the spelling error corrector to see what technique has been used. In addition, we present an evaluation of some existing medical spell correctors to test if they meet users' needs. We describe an ontology-based, spell checker architecture that would assist users with medical information retrieval.

Review of Spell Check Approaches

The literature on misspelling is diverse and much of it is based on studies of children. However, very little has been studied on adults. Review of papers show that spelling mistakes in adults are predictable (Yannakoudakis & Fawthrop, 1983). This section describes the types of spelling errors and the existing approaches to their correction.

General Pattern of Spelling Errors

A natural language such as English is not static. The English language is one of the most difficult languages to learn as it has the following properties:

- It is not a phonetic language; there are many words that are not spelled the way they sound.

Related Content

Parametric Survival Modelling of Risk Factor of Tuberculosis Patients under DOTS Program at Hawassa Town, Ethiopia

Fikadu Zawdie Chere, Yohannes Yebabe Tesfayand Fikre Enquoselassie (2015). *International Journal of Biomedical and Clinical Engineering (pp. 1-17).*

www.irma-international.org/article/parametric-survival-modelling-of-risk-factor-of-tuberculosis-patients-under-dots-program-at-hawassa-town-ethiopia/136232

Development of a Classification Model for CD4 Count of HIV Patients Using Supervised Machine Learning Algorithms: A Comparative Analysis

Peter Adebayo Idowuand Jeremiah Ademola Balogun (2019). *Computational Models for Biomedical Reasoning and Problem Solving (pp. 149-176).*

www.irma-international.org/chapter/development-of-a-classification-model-for-cd4-count-of-hiv-patients-using-supervised-machine-learning-algorithms/227275

Searching Bioinformatics Information Strategies for Effective Use of Search Engine

Viveka Vardhan Jumpala (2018). *Biomedical Engineering: Concepts, Methodologies, Tools, and Applications (pp. 742-748).*

www.irma-international.org/chapter/searching-bioinformatics-information-strategies-for-effective-use-of-search-engine/186704

An Agent-Based Architecture to Ubiquitous Health

Daniel Ruiz Fernández (2010). *Ubiquitous Health and Medical Informatics: The Ubiquity 2.0 Trend and Beyond (pp. 213-232).*

www.irma-international.org/chapter/agent-based-architecture-ubiquitous-health/42935

Infant Cry Detection and Pain Scale Assessment: A Pilot Study

N. Sriraamand S. Tejaswini (2014). *International Journal of Biomedical and Clinical Engineering (pp. 42-51).*

www.irma-international.org/article/infant-cry-detection-and-pain-scale-assessment/115884