Wright State University

CORE Scholar

11-2014

An Analysis of Mayo Clinic Search Query Logs for Cardiovascular Diseases

Ashutosh Sopan Jadhav
Wright State University - Main Campus, jadhav.5@wright.edu

Amit P. Sheth
Wright State University - Main Campus, amit@sc.edu

Jyotishman Pathak

Follow this and additional works at: https://corescholar.libraries.wright.edu/knoesis

Part of the Bioinformatics Commons, Communication Technology and New Media Commons, Databases and Information Systems Commons, OS and Networks Commons, and the Science and Technology Studies Commons

An Analysis of Mayo Clinic Search Query Logs for Cardiovascular Diseases

Ashutosh Jadhav, MS[1], Amit Sheth, PhD[1], Jyotishman Pathak, PhD[2]
[1]Knoesis Center, Wright State University, Dayton, OH; [2]Mayo Clinic, Rochester, MN

Abstract

Increasingly, individuals are taking active participation in learning and managing their health by leveraging online resources. Understanding online health information searching behavior can help us to study what health topics users search for and how search queries are formulated. In this work, we analyzed 10 million cardiovascular diseases (CVD) related search queries from MayoClinic.com. We performed semantic analysis on the queries using UMLS MetaMap and analyzed structural and textual properties as well as linguistic characteristics of the queries.

Introduction

Since the early 2000's, Internet usage for health information searching has increased significantly. According to the latest 2013 Pew Survey, one in three American adults have gone online to find out information about a medical condition. According to Center for Disease Control and Prevention (CDC), in the United States, CVD is one of the most common chronic diseases and the leading cause of death (1 in every 4 deaths) for both men and women. Prior studies have shown that online resources are 'significant information supplement' for the patients with chronic conditions. One of the most common ways to seek online health information is via Web search engines, such as Google. Therefore, studying search queries can help us to understand Online Health Information Seekers' (OHIS) "information needs" and how do they formulate search queries, which in turn, will empower us with knowledge to improve the health search experience, as well as to develop more advanced next-generation knowledge and content delivery systems. Although chronic diseases affect large population, very few studies have investigated online health information searching for chronic diseases and especially for CVD. We address this knowledge gap in the community by analyzing CVD related search queries. Some of the potential beneficiaries of this work are Web search engines and health websites.

Methods

In this study, we collected 10 million CVD-related search queries that direct users from Web search engines to the Mayo Clinic's consumer health information portal (MayoClinic.com). We performed following analysis on the CVD related search queries: 1) Top search queries associated with CVD 2) categorization of the queries into health categories using UMLS Metamap based on UMLS concepts and semantic types 3) Structural analysis: length of the search queries, usage of search query operators and special characters in the search queries 4) types of search queries (keyword based, Wh and Yes/No questions), misspellings in the queries, linguistic structure of the search queries.

Results and Discussion

Most of the top CVD queries are related to major CVD diseases and blood pressure (high/low). Top searched health categories for CVD are 'Diseases and Conditions' and 'Vital Signs'. Even though CVD prevention is possible, very few OHIS search for prevention while most of them search for symptoms and post diseases information (Living with, Diet, Treatment, Drugs). Average length of CVD search query (3.88 words and 22.22 characters) is longer than that of general search query, which implies that OHIS describe health information need in more details. Usage of search query operator (4%) is limited and variation of 'AND' (AND, &, +) is used more often (95%) followed by 'OR'. Only 3.2% of the search queries contain at least one spelling mistake. OHIS predominantly formulate search queries using keywords followed by Wh-Questions and Yes/No Questions. Almost all CVD search queries have at least one noun.

Conclusion

We found using Metamap and UMLS concepts/semantic type is very good approach for categorization of health related search queries into health categories. This study extends our knowledge about online health information searching behavior, and provides useful insights for Web search engines, health-centric websites, healthcare providers and healthcare-centric application developers.