Workshop on Health Search and Discovery: Helping Users and Advancing Medicine

Ryen W. White¹, Elad Yom-Tov¹, Eric Horvitz¹, Eugene Agichtein², William Hersh³
¹Microsoft Research, ²Emory University, ³Oregon Health & Science University
{ryenw,eladyt,horvitz}@microsoft.com, eugene@mathcs.emory.edu, hersh@ohsu.edu

ABSTRACT
This workshop brings together researchers and practitioners from industry and academia to discuss search and discovery in the medical domain. The event focuses on ways to make medical and health information more accessible to laypeople (including enhancements to ranking algorithms and search interfaces), and how we can discover new medical facts and phenomena from information sought online, as evidenced in query streams and other sources such as social media. This domain also offers many opportunities for applications that monitor and improve quality of life of those affected by medical conditions, by providing tools to support their health-related information behavior.

Categories and Subject Descriptors
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—search process, selection process; H.2.8 [Database Management]: Database applications—data mining.

Keywords
Health seeking; Medical discovery; Data mining; Social media.

1. BACKGROUND AND MOTIVATION
Health-related information is one of the most widely searched domains on the Web. A recent survey by the Pew Internet and American Life Project found that 59% of U.S. adults have looked online for health information in the past year, with 35% of respondents attempting to diagnose a medical condition online [1]. Health information seeking is conducted using both search systems such as Internet search engines and via social media (e.g., advice from friends). Research has indicated that cyberspace can be a dangerous place for those with no medical training. Web search engines can surface alarming content that can cause unwarranted anxiety [4].

Such challenges necessitate new information retrieval (IR) methods—both backend and frontend—to significantly improve the search experience for health information seeking. These methods can, for example, translate layman and technical terminology, and consider domain knowledge and interpretability and health-related reliability or authority during ranking or crawling [3]. This domain also presents unprecedented opportunities for the development of applications that monitor and improve the quality of life of people affected by a variety of medical conditions via tools to support their health-oriented information behavior. Mining aspects of that behavior (e.g., queries and social media interactions) with consent in the aggregate across many users has potential to assist in medical discoveries [5] and enhance public health monitoring [2].

The monitoring and use of health-related online behavior also brings into sharper focus important tradeoffs between privacy and benefit, which need to be explored in more detail. We envisage that the expertise of the SIGIR community will provide helpful insights on matters of privacy, as well as retrieval, personalization, expertise modeling, search interface design, data mining, and others; all of which are critical to enabling advances this area. Through presentations and discussions at the workshop we hope to outline how the IR community can play a central role the future of health search and discovery, and bring benefit to health seekers everywhere.

2. WORKSHOP OVERVIEW
The event comprises discussions (including brainstorming and breakouts), keynotes, and other presentations on papers submitted in response to an open call. The themes of the workshop include:

- Predictive analysis from large-scale data such as behavioral logs, microblogging, and other social media;
- Mining large-scale content (e.g., logs, Web crawls) for insights and links between conditions, drugs, and cohorts;
- Establishing the reliability of online health content;
- Diagnostic search by patients and health practitioners and its cognitive impact;
- Modeling the effect of domain expertise on health seeking;
- Query formulation for medical retrieval (including query expansion, UMLS, concept-hierarchies, etc.);
- Visualization and exploration of medical information;
- Privacy issues on the storage and retrieval of medical content;
- Personalization of short- and long-term health interests;
- Platforms for storing and maintaining health and fitness information (e.g., HealthVault), and;
- Social networks for sharing and monitoring health-related content (e.g., PatientsLikeMe).

Discussions at the event will target many of these topics. One area of particular interest is challenges in improving health search for laypeople. Websites such as WebMD.com, eHealthme.com, and PatientsLikeMe.com provide medical information or peer support to health seekers. We will discuss improvements to health search via, for example, synthesizing health content from many sources, and considering base rates and source reliability in result ranking.

REFERENCES