Medical Information Search Workshop

SIGIR 2016 Workshop Proposal

Steven Bedrick
Oregon Health State University
US
bedricks@ohsu.edu

Lorraine Goeuriot
LIG - UniversitÃ Ì grenoble Alpes
France
lorraine.goeuriot@imag.fr

Anastasia Krithara
NCSR Demokritos
Greece
akrithara@iit.demokritos.gr

Henning Mueller
HES-SO
Switzerland
henning.mueller@hevs.ch

George Paliouras
NCSR Demokritos
Greece
paliourg@iit.demokritos.gr

1. MOTIVATION FOR THE WORKSHOP AND APPROPRIATENESS TO SIGIR

1.1 Motivation

Medical information search refers to methodologies and technologies that seek to improve access to medical information archives via a process of information retrieval (IR). Such information is now potentially accessible from many sources including the general web, social media, journal articles, and hospital records. Health-related content is one of the most searched-for topics on the internet, and as such this is an important domain for research in information retrieval.

Medical information is of interest to a wide variety of users, including patients and their families, researchers, general practitioners and clinicians, and clinicians with specific expertise such as radiologists. There are several dedicated services that seek to make this information more easily accessible, such as Health on the Net's medical search systems for the general public and medical practitioners[1]. Despite the popularity of the medical domain for users of search engines, and current interest in this topic within the information retrieval research community, development of search and access technologies remains particularly challenging. One of the central challenges in medical information search is diversity of the users of these services. In particular, they will have varying categories of information needs, varying levels of medical knowledge, and varying language skills. These challenges can be summarized as follows:

1. Varying information needs: While a patient with a recently diagnosed condition will generally benefit most from simple or introductory information on the disease and its treatment, a patient living with or managing a condition over a longer term will generally be looking for more advanced information, or perhaps support groups and forums. In a similar way, a general practitioner might require basic information quickly while advising a patient, but more detailed information if deciding a course of treatment, and a specialist clinician might look for an exhaustive list of similar cases or research papers relating to the condition of a patient that they are currently seeking to advise. Understanding various types of users and their information needs is one of the cornerstones of medical information search, while adapting information retrieval to best address these needs to develop effective, potentially personalized systems is one of its greatest challenges.

2. Varying medical knowledge: The different categories of users of medical information search systems have different levels of medical knowledge, and indeed the medical knowledge of different individuals within a category can also vary greatly. This affects the way in which individuals pose search queries to systems and also the level of complexity of information which should be returned to them or the type of support in understanding / disambiguating returned material which will be required.

3. Varying language skills: Given that much of medical content is written in the English language, research to date in medical information search has predominantly focused on monolingual English retrieval. However, given the large number of non-English speakers on the Internet and the lack of content in their native language, effective support for them to search the English sources is highly desirable. The Internet in particular has affected the patient-physician relationship and providing relevant, reliable information to patients in their own language is a key to alleviate such challenging situations and lower phenomenons such as cyberchondria.

In addition, the format, reliability, and quality of biomedical and medical information varies greatly. A single health record can contain clinical notes, technical pathology data, images, and patient-contributed histories, and may be linked by a physician to research papers. The importance of health and medical topics and their impact on people's everyday lives makes the need for retrieval of accurate and reliable information especially important. Determining the likely reliability of available information is challenging.

Finally, as with information retrieval in general, the evaluation of medical search tools is vital and challenging. For example, there are no established or standardized baselines or evaluation metrics, and limited availability of test collections. Further discussion and progression on this topic

[1] http://www.hon.ch/

would be beneficial to the community.

1.2 Appropriateness to SIGIR

Information retrieval and related technologies are key to biomedical and medical information search. As such, papers on the topic are regularly published at SIGIR and other leading information retrieval conferences. However, as discussed, there are challenges hindering research progress. The current interest in medical information retrieval, as evidenced for example by activities such as the recent TREC Clinical Decision Support and Medical records tasks and the current CLEFeHealth tasks, means that hosting this workshop at SIGIR 2016 is extremely timely. A similar workshop was organized at SIGIR in 2014, and proved to be very successful. After the workshop, a special edition of the Journal of Information Retrieval (to appear) was published, with papers from workshop attendees, as well as other members of the community. We aim to bring together experts and active members of the medical information search community, and to give a unified understanding of the state-of-the-art and to identify specific research challenges to be explored by the community. We note that there is significant local interest in this problem area.

2. THEME AND PURPOSE OF THE WORKSHOP

The objective of this workshop is to provide a forum to enable the progression of research in medical information retrieval to provide enhanced search services for all users with interest in medical information search.

This workshop aims to bring together researchers interested in medical information search with the goal of identifying specific research challenges that need to be addressed to advance the state-of-the-art and to foster interdisciplinary collaborations towards the meeting of these challenges. To enable this, we will encourage participation from researchers in all fields related to medical information search including mainstream information retrieval, but also natural language processing, multilingual text processing, and medical image analysis.

Topics of interest include but are not limited to:

- Users and information needs
- Semantics and NLP for medical IR
- Reliability and trust in medical IR
- Personalised search
- Evaluation of medical IR
- Multilingual questions in medical IR
- Multimedia technologies in medical IR
- The role of social media in medical IR

3. FORMAT AND SCHEDULE

Our aim is to create a dynamic, interactive workshop structured to encourage discussion and active collaboration among attendees. Potential participants will be invited to submit papers for presentation at the workshop. A small group of carefully selected papers will be chosen for short oral presentation, while the remainder will be presented during a poster session. In addition to presentation of submitted work, we plan to invite a leading researcher in this area to deliver a keynote. Our first choice as invited keynote is William Hersh, Professor and Chair of the Department of Medical Informatics & Clinical Epidemiology in the School of Medicine at Oregon Health & Science University (OHSU) in Portland, Oregon, USA. Professor Hersh is a regular attendee at SIGIR, and one of the recognised leaders in the field of medical IR. A main focus of the workshop will be on discussion groups placing participants into groups with similar interests, and panel sessions enabling discussion of topics of general interest to the workshop. A final discussion session will close the event. This will seek to identify specific actions necessary to advance future research and development.

Tentative schedule of events: .

09:00 - 09:45 Welcome, workshop introduction, and attendee introductions
09:45 - 10:30 Keynote
10:30 - 11:00 Break
11:00 - 11:30 Paper session (three 10 minute slots)
11.30 - 11:45 Brainstorm several topics for breakout group discussion
11:45 - 12:30 Break into breakout groups and begin discussion
12:30 - 13:30 Lunch (breakout groups will be encouraged to discuss over lunch)
13:30 - 14:15 Breakout groups report back, with group discussion
14:15 - 15:00 Panel session
15:00 - 15:15 Poster boaster session
15.15 - 15:30 Poster session
15:30 - 16:00 Break (poster session continues during break)
16.00 - 16:15 Poster session (wrapup)
16:15 - 17.15 Round table discussion session
17:15 - 18:00 Wrap-up and roadmap of future directions

4. SELECTION PROCESS FOR PARTICIPANTS AND/OR PRESENTERS

The workshop will call for short papers (4 pages) and short position papers (2 pages). To establish the quality of submitted papers, we have brought together a group of researchers who have actively published on diverse themes within this topic and who are well known to the IR community to form the programme committee. All submissions will be blind reviewed by three PC members.

5. PAST ATTENDANCE AND OUTCOMES

A similar workshop was held at SIGIR 2014. Around twenty participants attended the workshop throughout the day, which included presentations, as well as a keynote talk and a discussion. Twenty papers were submitted, out of which 8 were selected for publication. After this workshop, organizers edited a special edition of the Journal of Information Retrieval (to appear). Eighteen papers were submitted and eight accepted, written by workshop attendees and other research groups.

We show in Section 8 that the interest in medical and health related data mining and information search is grow-

ing, with many conferences, workshops and evaluation challenges. Hence a second edition of this workshop would be of great interest. Moreover, holding it in Europe will allow groups who could not travel to Australia to participate and thus address the very active European medical retrieval communty.

6. LIST OF ORGANISERS

Dr Steven Bedrick.

Steven Bedrick is an assistant professor in Oregon Health & Science University's Center for Spoken Language Understanding. He obtained his PhD in biomedical informatics at Oregon Health & Science University, with a dissertation on cross-language medical information retrieval, and then worked as a postdoctoral researcher on text normalization in informal genre text. For several years, he was one of the organizers of the ImageCLEF medical image retrieval task, and in recent years has co-led the relevance judgment process for the TREC Medical and Clinical Decision Support tracks. He is currently involved in several different NIH research projects on NLP applications to assistive technology and neuropsychological assessment.

Dr Lorraine Goeuriot.

Lorraine Goeuriot is an associate professor in Université Grenoble Alpes. She obtained her Master in computer science and a PhD in computational linguistics on medical data in the University of Nantes, France. She worked as a post-doctoral researcher in Nanyang Technological University, Singapore, on medical opinion mining and in Dublin City University on medical information processing and retrieval. She is co-chair of the CLEF eHealth 2014 and 2015 evaluation lab, and has been co-leading the information retrieval task in 2013 and 2014. She was publication co-chair for SIGIR 2013 and COLING 2014. She has been involved in two national French research projects and currently in EU project Khresmoi, on medical information access. She is reviewing papers for several medical informatics workshops and journals.

Dr Anastasia Krithara.

Anastasia Krithara has been a post-doctoral researcher in the Institute of Informatics and Telecommunications at NCSR "Demokritos" since October 2008, where she is involved in national and international projects. Before, she was a research engineer in Xerox Research Centre Europe, in Grenoble, France, where she carried out research in the area of machine learning and more precisely in semi-supervised and active learning for text classification tasks. She holds a BSc in Computer Science from Athens University of Economics and Business (AUEB), an MSc in Machine Learning and Data Mining from University of Bristol and a PhD in Machine Learning from Pierre and Marie Curie University (Paris VI). Her research interests include Machine Learning, Information Retrieval, Knowledge Engineering and Natural Language Processing. She is program committee member of several international conferences and workshops and her work has been published in international journals, conferences and books. She has co-organised the International Research centered Summer Schools (IRSS-2013 and IRSS-2014), as well as the BioASQ challenges.

Pr Henning Müller.

Henning Müller obtained his Masters degree in medical informatics from Heidelberg University, Germany, in 1997 and his PhD on multimedia information retrieval from the University of Geneva, Switzerland in 2002. During this time he also worked for Daimler Benz research and technology North America in Portland, Oregon, USA, and at Monash University in Melbourne, Australia. After his PhD Henning has worked in medical informatics at the University and University hospitals of Geneva, Switzerland, where he finished his habilitation in 2008. Since 2007 he has been a professor in computer science at the University of Applied Sciences Western Switzerland in Sierre, Switzerland. Henning has lead the ImageCLEF benchmark on multilingual and multimodal information retrieval for ten years. He has published over 400 scientific articles on visual information analysis & retrieval and received several best paper/oster prices. He has participated in several EU projects, has initiated several national projects and coordinated the Khresmoi project on medical retrieval. He is currently on a sabbatical at the Martinos Center, Harvard Medical School in Boston, MA, USA.

Dr. George Paliouras.

George Paliouras is a senior researcher and head of the Intelligent Information Systems division of the Institute of Informatics and Telecommunications at NCSR âĂIJDemokritosâĂİ. He holds a PhD from the University of Manchester, UK, on machine learning. He has performed basic and applied research in artificial intelligence for the last 18 years and he has been publishing in all of the related conferences and journals. Among various contributions to the research community, he has served as board member in national and international scientific societies; he is serving in the editorial board of international journals and has chaired international conferences. He is involved in many European and national research projects and he has the role of scientific coordinator in some of them. He has taught postgraduate courses and has co-founded NCSR's spin-off company i-sieve technologies Ltd.

7. POTENTIAL PROGRAM COMMITTEE

A PC will be formed to review submissions to ensure that they provide high quality contribution to the workshop. The following researchers have been contacted. Some, as indicated, have already accepted our invitation.

- Peter Bruza, Queensland University of Technology
- Ben Carterette, University of Delaware (accept)
- Wendy Chapman, University of Utah
- Allan Hanbury, Vienna University of Technology (accept)
- David Hawking, Bing (Microsoft)
- William Hersh, Oregon Health and Science University (accept)
- Jung-Jae Kim, Nanyang Technological University (accept)
- Iadh Ounis, University of Glasgow (accept)

- Patrick Ruch, HES-SO Geneve (accept)
- Stefan Schulz, Medical University Graz
- Karin Verspoor, NICTA
- Ellen Voorhees, NIST (accept)
- Ryen White, Microsoft Research (accept)
- Elad Yom-Tov, Microsoft Research (accept)
- Pierre Zweigenbaum, LIMSI (accept)

8. RELATED WORKSHOPS

In recent years, workshops have been organised on medical natural language processing and text mining. BioNLP for instance is an ACL associated workshop that has been organised since 2002. This workshop targets foundational research in language processing (NLP) for the biological and medical domains. A shared task track is collocated with the workshop and organises text mining and fine-grained information tasks (such as BioASQ, clinical question answering, genia event extraction for NfkB knowledge base construction and cancer genetics discovery). The Louhi workshop offers another forum for researchers to explore health text mining and information analysis. This year it is organised at EACL and previously it was collocated with other events.

While these two workshops are centred on the technologies and methodologies for medical information extraction and NLP, a workshop focused on building and evaluating resources for biomedical text mining has taken place for the past four years at the Language Resources and Evaluation Conference (LREC). This workshop targets papers describing resources and issues relating to their usability.

The IR community has tackled issues related to the medical domain via shared tasks over several years: TREC Genomics track (2003-2007), TREC Medical records track (2011-2012), TREC Clinical Decision Support (2014-2015), ImageCLEF (2003-2016), CLEF eHealth (2013-2016), BioASQ (2013-2016). While these tracks mainly focus on medical information retrieval, they are each centred on one task only and offer little space for wider exploration.

A few workshops have focused on Information retrieval in the recent years. In SIGIR, the workshop on health search and discovery (helping users and advancing medicine), was organised in 2013. It sought to investigate ways to make medical and health information more accessible to laypeople and ways to discover new medical facts and phenomena from information sought online. This exploratory workshop proved that the IR community has many links with the medical domains and highlighted many directions to explore. One of the key directions for further exploration was information retrieval techniques for the medical space. As a next step in progression, our proposal is focusing on this direction: information retrieval, with a primary focus on its multimodal and multilingual aspects. This year, the Multi-Perspective Consumer Health Search (MPCHS) Workshop, organized at WWW'16 seeks to investigate methods to handle consumer queries without any simple answer.

This Medical Information Retrieval workshop was held first in SIGIR in 2014. Given the amount of submissions received, the success of the JIR special edition, and the growing interest of researchers into this domain, a second edition of this workshop should be interesting to the community.