A CASE OF TASK-BASED SEARCH IN THE MEDICAL DOMAIN
RADO DRAGUSIN, PAULA PETCU, CHRISTINA LIOMA, OLE WINTHER

dragusin@diku.dk paula.petcu@findwise.com lioma@gmail.com owi@imm.dtu.dk

Computer Science Department, University of Copenhagen, Denmark
FINDWISE Findwise ApS, Denmark
DTU Bioinformatics, Technical University of Denmark

1 RARE DISEASE DEFINITION

- affecting <1 in 2,000 EU citizens

- ~30 million rare disease patients in EU

- 40% are misdiagnosed

- 25% suffer diagnostic delays of 5-30 yr

- 5,000 – 8,000 known rare diseases in EU

- 80% of genetic origin

rare diseases are often difficult to diagnose:

- large number & variation (1 in 10,000)
- low prevalence (1 in 10,000)
- non-specific symptoms

Although clinicians will encounter many rare disease patients throughout their career, they will have little practical experience with specific diseases since the likelihood of encountering them more than once is small.

2 TASK-BASED SEARCH FOR RARE DISEASES

- when confronted with difficult cases, clinicians

traditional use: { books, journals, colleagues }

increasingly use: {Google, PubMed, Orphanet }

3 ZEBRA SEARCH ENGINE

task integration

- easy-to-use interface
- generate diagnostic hypotheses
- takes patient data as free text

index

- 31,114 medical documents
- 19 online medical sources

evaluation

on 56 difficult real-life medical cases

- Zebra:
  - 67.9% of test cases return the correct disease (MRR 0.385)

- Google:
  - 37.5% of test cases return the correct disease (MRR 0.206)

For some of the test cases we observed the retrieval of multiple articles describing the same diseases.

A task-based search engine would better fit the clinician’s task-specific needs, being tailored for the workflow and time restrictions of the diagnostic process.

4 DOCUMENT CLUSTERING BASED ON MEDICAL CONCEPTS

map documents to diseases

- most indexed documents cover one disease or group of diseases
- map document titles to diseases (using terms from UMLS Metathesaurus)

document ↔ disease

- cluster documents by diseases
- search by disease name
- rank diseases, not documents

5 DISEASE RANKING

diagnosing difficult cases

- iterative process
- clinicians interested in disease hypotheses

ranking diseases

- disease score based on supporting documents
- a more natural framework for selecting hypotheses

Workshop on Task-Based and Aggregated Search (TBAS), 34th European Conference on Information Retrieval (ECIR), 1-5 April 2012, Barcelona, Spain