Diplomarbeitspräsentation

FAKULTÄT FÜR !NFORMATIK

Faculty of Informatics

Using health statistics to improve medical and health search

Masterstudium:
Medizinische Informatik

Tawan Sierek

Technische Universität Wien
Institut für Softwaretechnik und Interaktive Systeme
Information & Software Engineering Group
Betreuer: Privatdoz. Dr. Allan Hanbury

Motivation

We assume: frequencies of diseases can be used as a relevance signal within health search. Documents, which cover more frequent diseases, are more relevant than documents, which cover rare diseases. **Research Questions:**

- *Can epidemiological data improve ranking in health search?*
- *Which epidemiological data sources are suitable?*
- *Can ranking be improved by adapting to a patient's sex and age?*

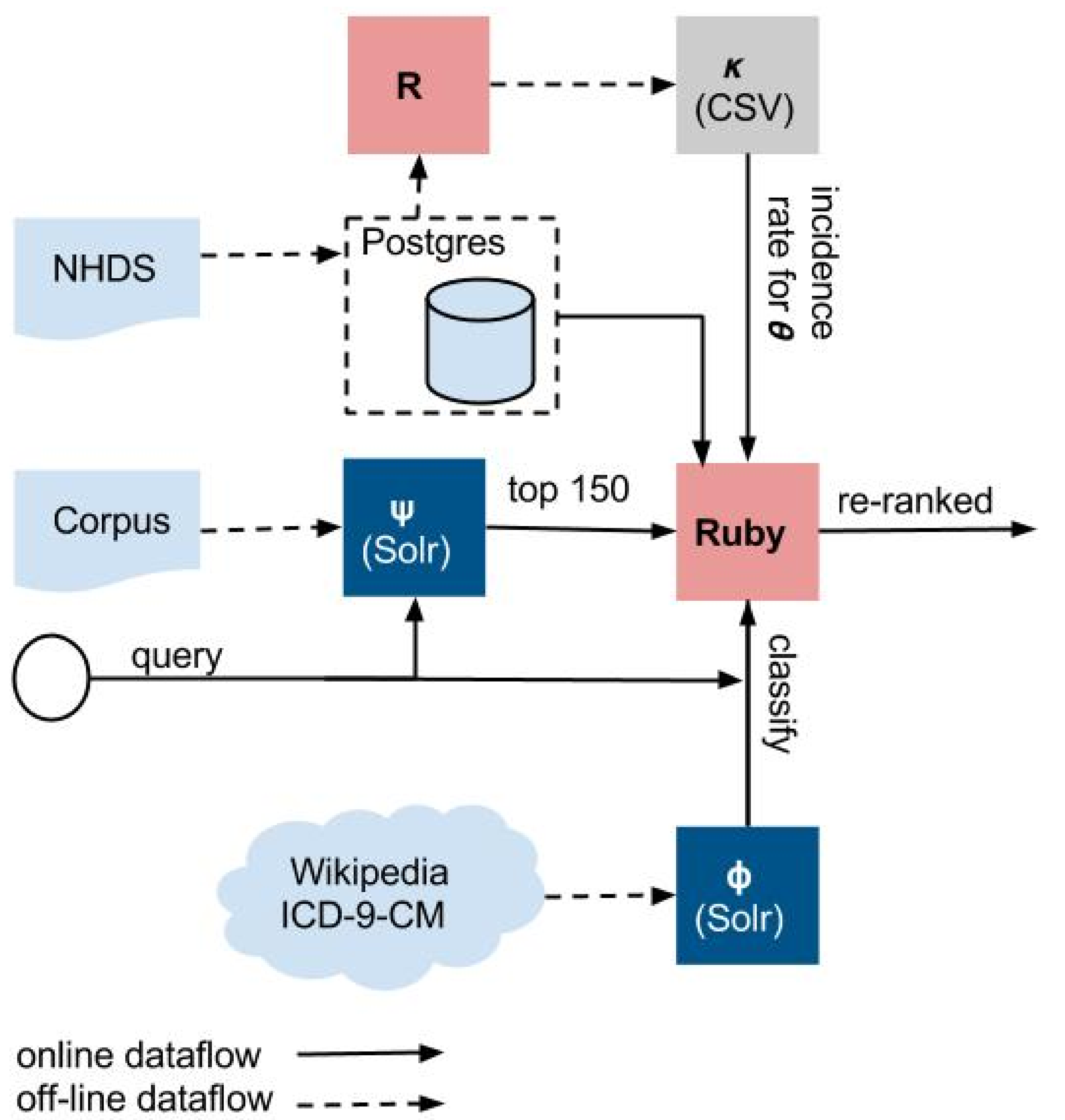
Probabilistic Health Search Model (PHS)

$$\hat{p}(R = 1 \mid \theta, d, q) = \psi(d, q) \sum_{C} \phi(C, d) \frac{\kappa(C, \theta)\phi(C, q)}{\sum_{C'} \kappa(C', \theta)\phi(C', q)}.$$

R	takes on $\mathbf{1}$ if d is relevant to q, otherwise $\mathbf{0}$.
θ	the patient profile.
d	the document.
q	the query.
ψ	patient and disease independent relevance signal.
C	a disease.
κ	maps to the incidence rate.
ϕ	text classifier, assigns text to disease.

The model was inspired by Sontag et al. and their probabilistic models to personalize general web search [1].

Reference Implementation



We use the US National Hospital Discharge Survey (NHDS) 2007 to estimate incidence rates. The diagnoses are encoded in ICD-9-CM. Our text classifier uses Wikipedia articles, which describe ICD-9-CM codes, as training data.

The text classifier was inspired by the work of Trieschnigg et al. [2].

Evaluation

We evaluated the reference implementation on two IR tracks:

- **TREC Clinical Decision Support Track 2014 (CDS)**
  - 733,138 articles of the PubMed Central Archive.
  - 30 topics, 3 categories: diagnosis, test, treatment.
  - goal: relevant information for physicians when presented with a medical case narrative.
- **ShARe/CLEF eHealth Evaluation Lab 2014 Task 3 (CLEF)**
  - 1,104,337 documents about health topics, from a large scale web crawl.
  - 50 topics
  - goal: retrieve information for patients, who want to understand their health information.

Results

The results, shown here, were produced with runs that were not adapted to a patient's age and sex. Adapted runs did not show any improvements.

Measure	Baseline	PHS
MAP	0.1208	**0.1222**
NDCG@5	0.3188	**0.3308**
NDCG@10	0.2732	**0.2885**
P@5	0.3667	**0.3733**
P@10	0.3033	**0.3167**

(a) CDS Track

Measure	Baseline	PHS
MAP	0.3951	**0.396**
NDCG@5	**0.7261**	0.7164
NDCG@10	**0.7123**	0.706
P@5	**0.752**	0.74
P@10	0.718	**0.72**

(b) CLEF Track

Table : Results: (a) on the CDS track the PHS model shows minor improvements in regard to all measures, (2) on the CLEF track the PHS model did not show any improvements.

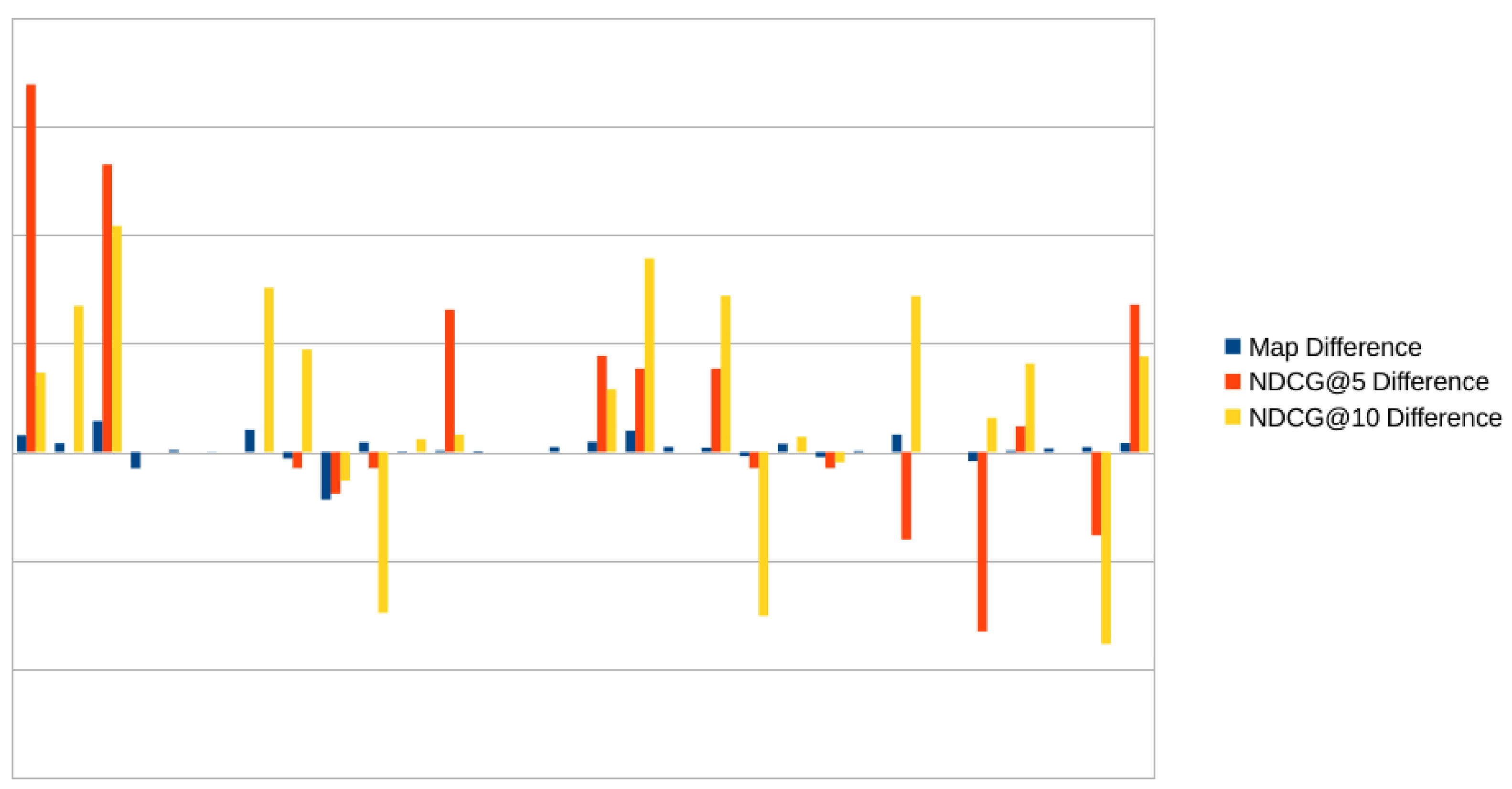


Figure : Single topic differences of PHS and baseline on the CDS track (PHS - baseline).

Improvements are too small and **not** statistically significant.

Conclusions:

- Results are confounded by performance of text classifier, which we can not evaluate, ICD-9-CM training data is not really available.
- Incorporating a patient profile (age, sex) seems to make no difference.
- We suggest further experiments with manually annotated documents.

References:

[1] D. Sontag, K. Collins-Thompson, P. N. Bennett, R. W. White, S. Dumais, and B. Billerbeck. Probabilistic Models for Personalizing Web Search. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, pages 433–442, New York, NY, USA, 2012. ACM.

[2] D. Trieschnigg, P. Pezik, V. Lee, F. de Jong, W. Kraaij, and D. Rebholz-Schuhmann. MeSH Up: effective MeSH text classification for improved document retrieval. *Bioinformatics*, 25(11):1412–1418, June 2009.

Kontakt: tawan@sierek.at