



# ZEBRA: SEARCHING FOR RARE DISEASES

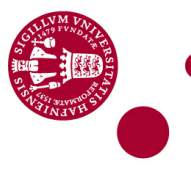
## A CASE OF TASK-BASED SEARCH IN THE MEDICAL DOMAIN

RADU DRAGUSIN<sup>1</sup>, PAULA PETCU<sup>2</sup>, CHRISTINA LIOMA<sup>1</sup>, OLE WINTHER<sup>3</sup>

dragusin@diku.dk


lioma@imm.dtu.dk

owj@imm.dtu.dk



Computer Science Department,  
University of Copenhagen,  
Denmark

**FINDWISE**  
<sup>2</sup>Findwise Aps., Denmark



<sup>3</sup>DTU Informatics,  
Technical University  
of Denmark

**DTU**

## 1 RARE DISEASE DEFINITION

## 2 TASK-BASED SEARCH FOR RARE DISEASES

*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* (5 – 8 thousand)
- low prevalence* ( $< \frac{1}{2000}$  affected)
- 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.

*when confronted with difficult*

*cases, clinicians*

traditionally  
use:

*books  
journals  
colleagues*

increasingly  
use:

*Google  
PubMed  
Orphanet*

*information retrieval systems*

- good at matching queries to large medical corpora*

*clinicians*

- good at filtering unsuitable results*

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.

## 3 ZEBRA SEARCH ENGINE

*task integration*

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

*index*

- 31,114* medical documents
- 10* 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*)

system	MRR	P@10	P@20	correct answers in top 10	correct answers in top 20
Zebra	0.385	0.125	0.089	35 (62.5%)	38 (67.9%)
Google	0.206	0.088	0.071	16 (28.6%)	18 (32.1%)

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

...

ZebraHomeAbout

girl, hypotonia, seizures, dehydration, polyphnea, acidosis, massive ketonuria, hyperammonemia

1Methylmalonic Aciduria, Cblb Type

23-@Methylcrotonyl-CoA Carboxylase 1 Deficiency

3Methylmalonic Aciduria, Cblb Type

4Propionic Acidemia

5Methylmalonic Aciduria Due To Methylmalonyl-CoA Mutase Deficiency

6Caniline Palmitoyltransferase 1 Deficiency

7Fatal infantile lactic acidosis with methylmalonic aciduria

1Methylmalonic acidemia

2PROPRIONIC ACIDEMIA

3Propionic Acidemia

4Propionic acidemia

5Acidemia, Propionic

Search for more like this...

1C0304341PYRUVATE CARBOXYLASE DEFICIENCY

2C14109143-@HYDROXY-3-METHYLGUTARYL-CoA LYASE DEFIC...

3C0208575ISOVALERIC ACIDEMIA

4C0208575ISOVALERIC ACIDEMIA

5C0208575ISOVALERIC ACIDEMIA

6C0208575ISOVALERIC ACIDEMIA

7C0208575ISOVALERIC ACIDEMIA

regular search

1girl, hypotonia, seizures, dehydration, polyphnea, acidosis, massive ketonuria, hyperammonemia

23-@Methylcrotonyl-CoA Carboxylase 1 Deficiency

3Methylmalonic Aciduria, Cblb Type

4Propionic Acidemia

5Methylmalonic Aciduria Due To Methylmalonyl-CoA Mutase Deficiency

6Caniline Palmitoyltransferase 1 Deficiency

7Fatal infantile lactic acidosis with methylmalonic aciduria

## 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

...

ZebraHomeAbout

girl, hypotonia, seizures, dehydration, polyphnea, acidosis, massive ketonuria, hyperammonemia

1Methylmalonic Aciduria, Cblb Type

2Methylmalonic Aciduria, Cblb Type

3Vitamin B12-unresponsive methylmalonic aciduria

43-@Methylcrotonyl-CoA Carboxylase 1 Deficiency

5Propionic Acidemia

6Propionic acidemia

7Acidemia, Propionic

8Methylmalonic Aciduria Due To Methylmalonyl-CoA Mutase Deficie...

1Propionic Acidemia; Propionyl-Coa Carboxylase Deficiency; Pcc Deficiency; Glycinemia, Ketotic; Hyperglycinemia With Ketoacidosis And Leukopenia, Ketotic Hyperglycinemia

http://www.ncbi.nlm.nih.gov/omim/606054

**clustering results**

girl, hypotonia, seizures, dehydration, polyphnea, acidosis, massive ketonuria, hyperammonemia

1girl, hypotonia, seizures, dehydration, polyphnea, acidosis, massive ketonuria, hyperammonemia

23-@Methylcrotonyl-CoA Carboxylase 1 Deficiency

3Methylmalonic Aciduria, Cblb Type

4Propionic Acidemia

5Propionic acidemia

6Acidemia, Propionic

7Methylmalonic Aciduria Due To Methylmalonyl-CoA Mutase Deficiency

ZebraHomeAbout

girl, hypotonia, seizures, dehydration, polyphnea, acidosis, massive ketonuria, hyperammonemia

1C0208583Methylmalonic acidemia

2C0208579PROPRIONIC ACIDEMIA

3Propionic Acidemia

4Propionic acidemia

5Acidemia, Propionic

UMLS concept name:  
**PROPRIONIC ACIDEMIA**

1C0304341PYRUVATE CARBOXYLASE DEFICIENCY

2C14109143-@HYDROXY-3-METHYLGUTARYL-CoA LYASE DEFIC...

3C0208575ISOVALERIC ACIDEMIA

4C0208575ISOVALERIC ACIDEMIA

5C0208575ISOVALERIC ACIDEMIA

6C0208575ISOVALERIC ACIDEMIA

7C0208575ISOVALERIC ACIDEMIA

Concept found in 3 of the retrieved documents.  
Concept found in 7 of the documents in our collection.  
Variant naming: PROPRIONIC ACIDEMIA.  
Concept ID: C0208579.

disease ranking

http://findzebra.com

## 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), 34<sup>th</sup> European Conference on Information Retrieval (ECIR), 1-5 April 2012, Barcelona, Spain

The author has requested enhancement of the downloaded file. All in-text references underlined in blue are linked to publications on ResearchGate.