Informationsextraktion aus medizinischen Case-Studies

Tamás Petrovics



Allan Hanbury





1. Introduction to XUND

2. The science behind DISCO

3. DISCO in action



Over **50% of people** don't understand basic health information...

and try to fill this knowledge gap by consulting **Dr.Google**.



What to do abou	it my headache?		ļ
Did you mean: brain tumor stroke brain abscess bacterial mening	itis		
	Google Search	I'm Feeling Lucky	



Our **digital health assistant empowers users** to make informed and better decisions



Understand

Providing fast and valid medical feedback on their symptoms in acute situations...

Prevent

and helping to track and analyze health over time to mitigate risks and increase quality of life.

Our key to success and market leadership is based on 3 core principles



Human-Al Interaction

Al-powered data mining of millions of peer reviewed sources to create scalable fact-based for medical team



Quality First Mindset

Standardised processes and systems to ensure highest quality standards are met for classification as a medical device



Engaging UX

Agile and intuitive user interface that efficiently navigates patients along their journey for maximized symptom prediction



DISCO is our AI solution developed together with TU Wien, that powers the Medical Engine behind XUND



1. Introduction to XUND

2. The science behind DISCO

3. DISCO in action



Output ranked list of symptoms:







M Informatics

Usual Approach: Keywords

Periodontitis associated with Chediak-Higashi syndrome in a young African American male.





Usual Approach: Keywords

Periodontitis associated with Chediak-Higashi syndrome in a young African American male.

formatics



Our approach: Extract the symptoms from the full text

<u>BMC Gastroenterol</u>. 2015; 15: 18. Published online 2015 Feb 13. doi: <u>10.1186/s12876-015-0242-1</u> PMCID: PMC4329676 PMID: <u>25888516</u>

Epidemiology of appendicitis and appendectomy for the low-income population in Taiwan, 2003–2011

Kai-Biao Lin,^{#1,5} Chien-Lung Chan,^{®#2,6} Nan-Ping Yang,^{#3,4} Robert K Lai,^{#5,6} Yuan-Hung Liu,^{5,6,7} Shun-Zhi Zhu,¹ and Ren-Hao Pan⁶

Author information
Article notes
Copyright and License information <u>Disclaimer</u>

This article has been cited by other articles in PMC.

Abstract

Go to: 🕑

Background

Although numerous epidemiological studies on appendicitis have been conducted worldwide, only a few studies have paid attention to the effect of socioeconomic status on appendicitis, particularly studies focusing on the low-income population (LIP).

Methods

We analyzed the epidemiological features of appendicitis in Taiwan using data from the National Health Insurance Research Database from 2003 to 2011. All cases diagnosed as appendicitis were enrolled.

Results

Between 2003 and 2011, 2,916 patients from the LIP and 209,206 patients from the normal population (NP) were diagnosed with appendicitis. Our finding revealed that the ratios of comorbidities, complicated appendicitis, and readmissions in LIP patients were slightly higher than those of NP patients. LIP patients were more likely to live in suburban and rural areas, and hence a higher proportion of them were hospitalized in a district or regional hospital compared with NP patients. The crucially finding was that the overall incidence ratios of appendicitis, acute appendicitis, and perforated appendicitis in the LIP were substantially higher than those in the NP (36.25%, 35.33%, and 37.28%, respectively). The mean LOS in LIP patients was longer than that of NP patients. The overall case-fatality ratio of appendectomy in the LIP was higher when compared with the NP (0.41% versus 0.12%, p < 0.05). We also observed that appendiciti was occurred frequently in male patients, with a higher incidence for those aged 15-29 years.

1,542,847 PubMed Central Articles (PMC)

- Free to use even for commercial purposes
- Each article includes: title, abstract, full text

Informatics

Preparation Phase:

Approach - Outline



Symptom Extraction Phase:



Symptom Ranking

The ranking function for symptoms is key

- Symptoms like *Pain* will occur quite frequently but aren't that useful
- Symptoms like Abdominal Pain will occur quite frequently in articles about Appendicitis, but occur less often for random articles



How well does it work?

- Dataset: Manual collection of 20 diseases and their symptoms based on high quality sources, e.g. textbooks
- 3 Medical Doctors annotate each symptom as primary or not
 - agreement: Overall Fleiss Kappa of 0.61

nformatics

Table 1: Overview of the DSR-collection. For each disease, we display the number of symptoms (#S), the number of primary symptoms (#P), and the Fleiss' interannotator agreement (κ).

Disease	#S	$\#\mathbf{P}$	κ	Disease	#S	$\#\mathbf{P}$	κ
Anorexia Nervosa	7	2	1.00	Influenza	11	2	0.57
Appendicitis	7	2	1.00	Measles	9	4	0.38
Asthma	9	4	0.76	Mental Depression	13	3	0.21
Bronchitis	9	1	0.71	Migraine Disorders	12	4	0.37
Cholecystitis	12	1	0.55	Myocardial Infarction	11	4	0.44
COPD	7	3	0.83	Periodontitis	3	4	0.46
Diabetes Mellitus	11	3	0.72	Pulmonary Embolism	13	2	0.83
Epididymitis	8	2	0.67	Sleep Apnea Syndromes	13	2	0.31
Erysipelas	$\overline{7}$	3	0.69	Tonsillitis	7	4	0.63
GERD	8	2	0.76	Trigeminal Neuralgia	3	3	0.28

Evaluation results on the dataset created

Table 2: Comparison of the disease-symptom extraction methods using our proposed DSR-collection. We show significant improvements with: a refers to EMBEDDING, b to KWD, and c to KWDLARGE (student's t-test: p < 0.01).

Method	nDCG@5	P@5	R@5	nDCG@10	P@10	R@10
Embedding	0.20	0.18	0.08	0.19	0.15	0.13
COOCCUR-KWD	0.24	0.22	0.09	0.20	0.15^{a}	0.11
COOCCUR-KWDLARGE	0.32^{a}	0.27	0.12^{a}	0.28^{ab}	0.19^{a}	0.17^{b}
CoOccur-FullText	0.41^{abc}	0.39 ^{<i>abc</i>}	0.17^{abc}	0.36^{abc}	0.27^{ac}	0.25^{ab}



Output ranked list of symptoms:









Outcomes of the Cooperation for the TU Wien

- Very interesting problem to solve
- Access to medical expertise
- Science
 - Part of a Dissertation
 - Master Thesis
 - Publications





1. Introduction to XUND

2. The science behind DISCO

3. DISCO in action



The DISCO Analysis is the first step in our medical research process and creates the **quantitative fact base** for our medical engine

- ★ Faster development of medical database
- ★ Less dependency on human capacities
- ★ Significant cost advantages without reduced quality of output
- ★ More solid clinical evaluation (CE-Mark)
- ★ New insights gained e.g. for COVID-19





With the help of DISCO, we substantially outperform the competition



Medical accuracy

Compared to 53% competitor average¹



Symptom prediction

Compared to 50% competitor average¹

01:51

Assessment speed

Compared to 02:10 competitor average¹



¹ Over 100 verified medical cases from peer reviewed sources were run through the leading digital health assistants on the market (Ada, Babylon, Healthtap, Symptomate and XUND) and more than 1000 symptom assessments were performed to analyze their performance under identical and objective circumstances