

Medical Information Extraction: from patient dossiers to structured forms

Evangelos Kanoulas (IVI, UvA), Lydia Mennes (CTcue), Christina Zavou (UvA)

partners



UNIVERSITEIT VAN AMSTERDAM



Hogeschool van Amsterdam

CWI

The Task

- Patients' dossiers – Isala Hospital, Zwolle

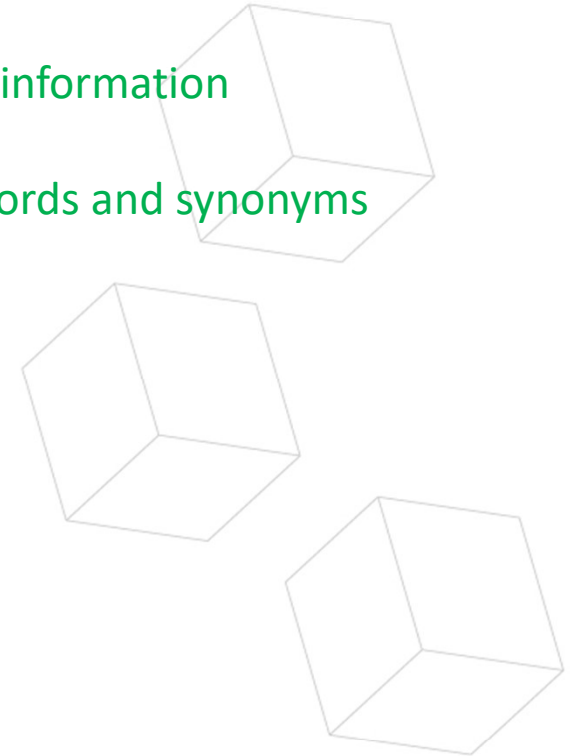
date	2012-06-01T00:00:00
type	None
description	[PATIENTID]/MS/AV NEWLINE NEWLINE BSN: [BSN]. NEWLINE NEWLINE Zwolle, 31-05-2013 ...met 23- 02-2011 op de afdeling Chirurgie. Diagnose(s): Rectaal carcinoom waarvoor bestraling Therapie: 18-02- 2011 OK: laparoscopische rectumamputatie NEWLINE NEWLINE Beloop: Geen complicaties. NEWLINE NEWLINE Pathologie: Rectumamputatie (status na korte radiotherapie): mucineus adenocarcinoom

care_request	None
care_provider_code	525729
treatment_code	None
end_date	2015-03-21T00:00:00
care_type	vervolg
dbc_status	C
diagnosis_code	132
specialty	Radiotherapie
start_date	2014-05-01T00:00:00
care_type_code	21
care_request_code	None
diagnosis	Rectale tumoren
specialty_code	361
327	None
treatment	None

| Approach



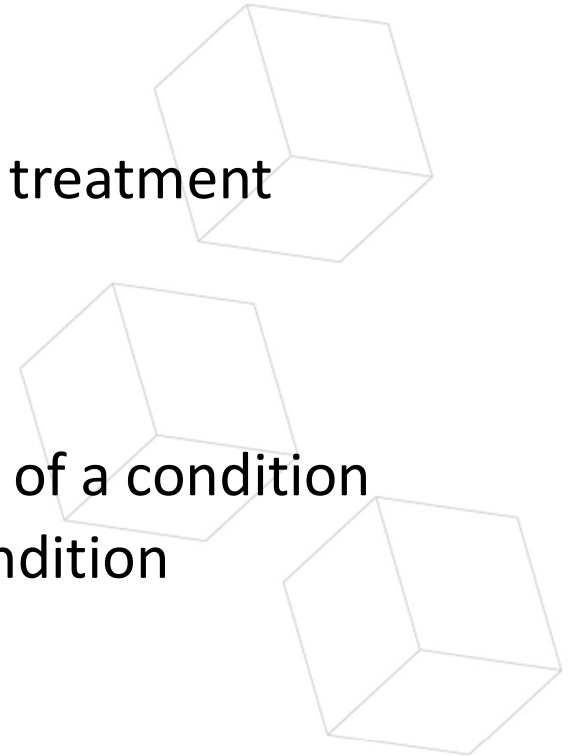
- Information Extraction
 - Find small pieces of text likely to contain important information
 - Restricted to few fields
 - Enriched the search queries with misspelled words and synonyms
- Deep Learning
 - Use supervision on top of IR



| Conclusion



- Hard task
 - Each field seems to require a different treatment
- Skewed data
 - Skewness mainly towards the absence of a condition
 - IR searches for the presence of the condition



| Future Work



- Supervision
 - Already represented text as presence/absence of a term, using the Boolean method to select which files to consider
 - Selection on the text to consider
 - Use raw features (i.e. words) in a deep learning technology

