Mining Patient Generated Data

"We'd like to start out being very involved with you but eventually be drawn away to much more interesting cases on Twitter."

The Rise of Social Media

Relationship with T4 to Diabetes 2

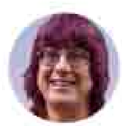

Shon 2 years ago 7 Replies


HealthUnlocked

Hi. I have Graves disease, 6 months ago I increased my thyroxine, and felt better for it. this took my T4 levels to the limit but my T3 is still midway, (apparently exceptable)
Two months ago I was diagnosed as borderline Diabetic 2. I am not convinced of this, as I feel the coincidence between me increasing my thyroxine from 150 to 175 per day to then suddenly becoming diabetic (bordeline) to easy. My theory is that by lifting my T4 level as high as it can go (safely) that this as increased my insulin or glucose levels. causing me to be showing as diabetic.
I there anyone out there that has any kno

sean

THYROXINE DIAGNOSED INSULIN GI
T3 DISEASE T4 DIABETES 2


JudyS 2 years ago

Here is another piece of the puzzle to confuse you - I recently had the genetic test for DI02 and found that I have the heterozygous genetic fault (which means I have a problem with converting T4 to T3) and the genetic counsellor told me there is a link between this fault and Type 2 diabetes, so it is indeed possible for you to have both.

It might be worth considering getting the DI02 test done yourself. Details are on the front page of the Thyroid UK site.

Reply Recommend (0)

Mining healthcare social media

- Extract
  - symptoms, problems
  - adverse events
  - treatments
  - life-style concepts
  - sentiment
  - psychological indicators
  - **quality of life**

Mining healthcare social media

Psychiatric disorders - Insomnia

When I’m on 60mg **prednisolone** I can’t sleep and want to eat 24/7. This is my last day on high dosage.□

Metabolism and nutrition disorders - Increased appetite

Mining ADRs and benefits of steroids

HealthUnlocked

Mining healthcare social media

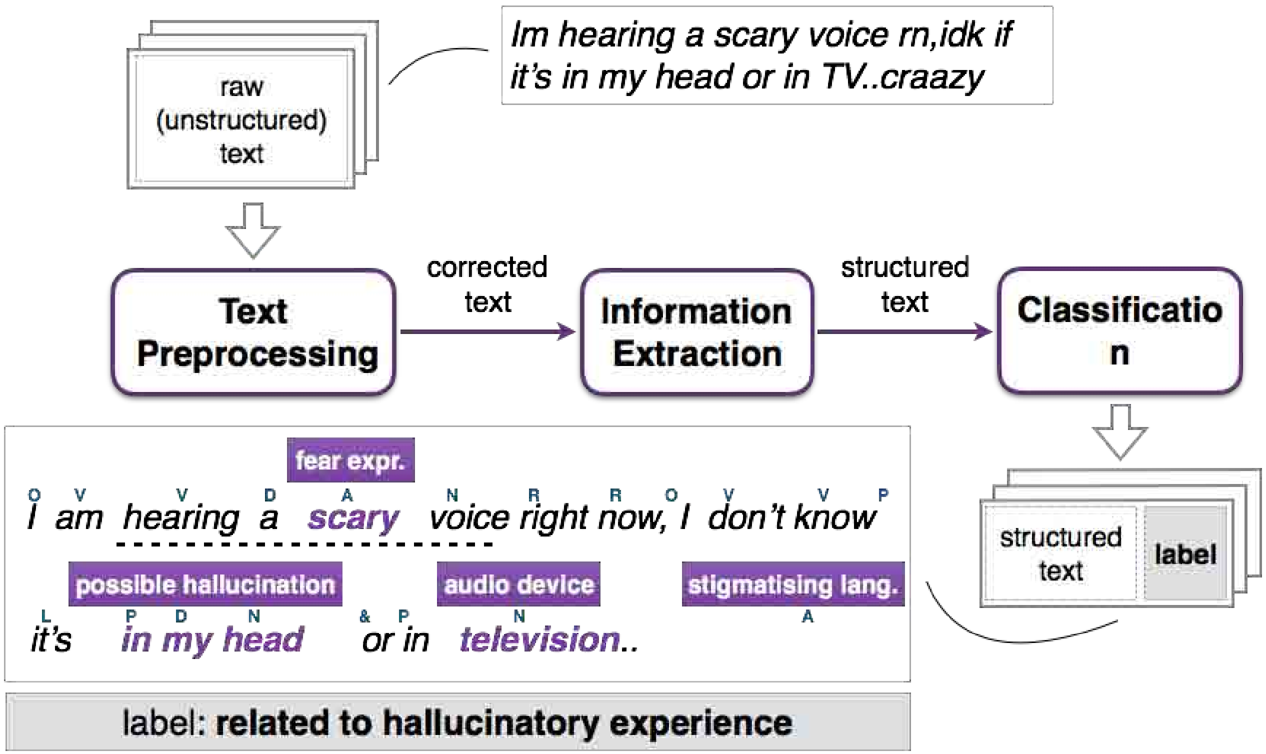
- Can we use Twitter to generate mineable datasets from unsolicited posts regarding risk factors for people with schizophrenia
  - e.g. auditory hallucinatory experiences
  - e.g. sleep-related issues
  - e.g. suicidal thoughts

> If hallucinating is thought of as hearing voices that are not actually real, then these painkillers are causing me to hallucinate like mad

> So I was convinced I was hearing stuff. It was so funny because the noise was coming from the kitchen but I thought I was hallucinating

Mining healthcare social media

Text classification pipeline

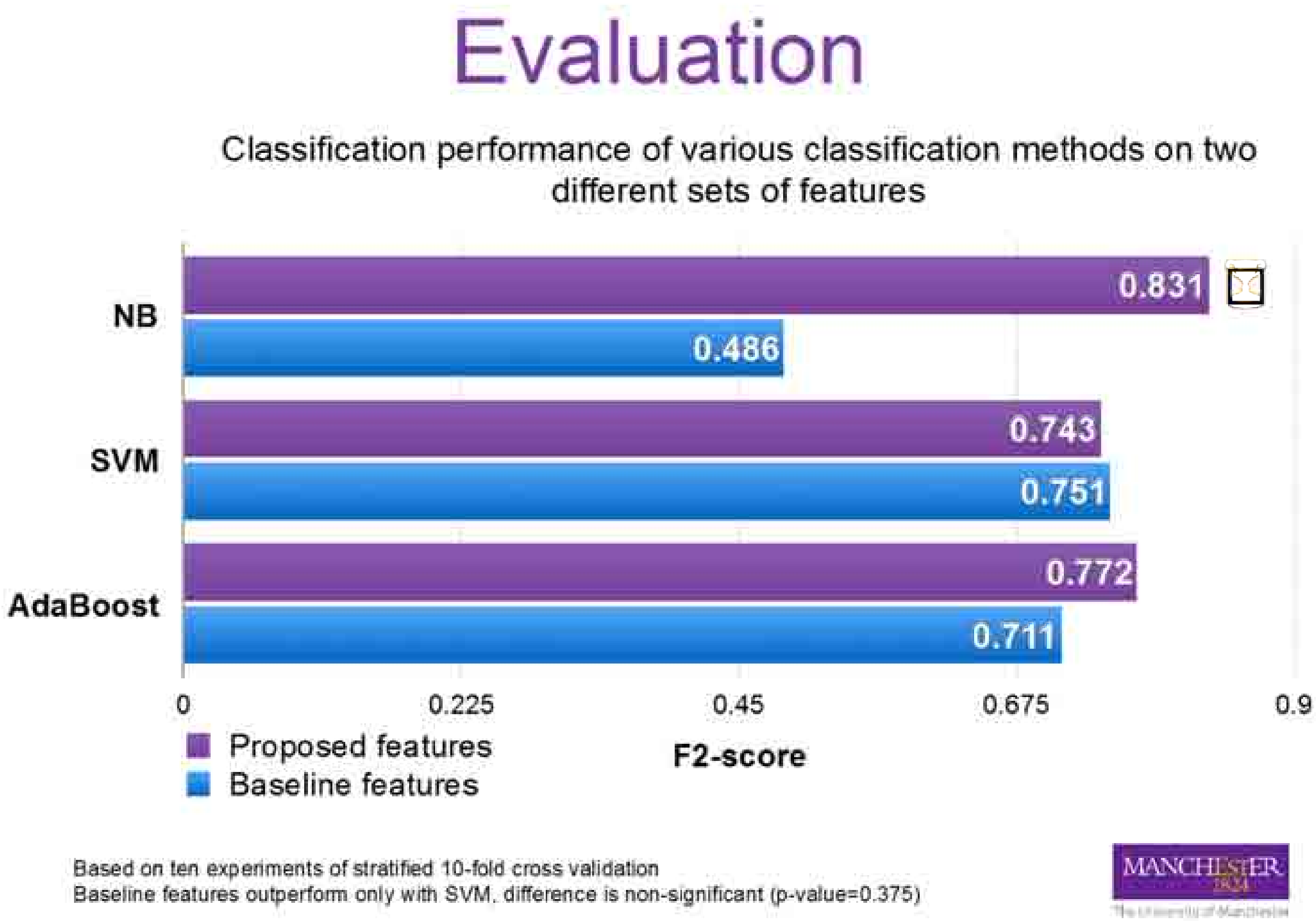


POS tagset from Gimpel et al. (2011): O - personal pronoun, V - verb, D - determiner, etc.

M Belousov et al.: *Mining Auditory Hallucinations from Unsolicited Twitter Posts.* Proceedings of LREC 2016 Workshop.

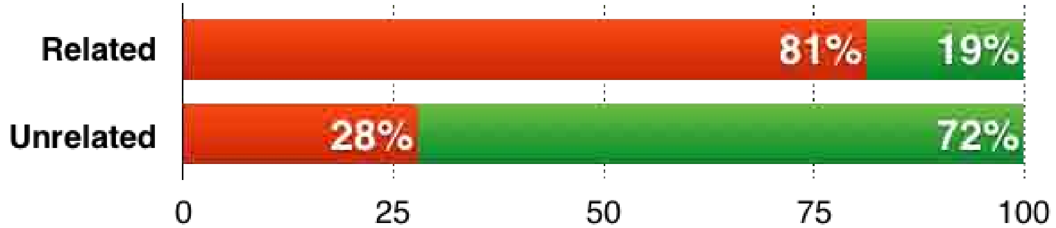
Mining healthcare social media

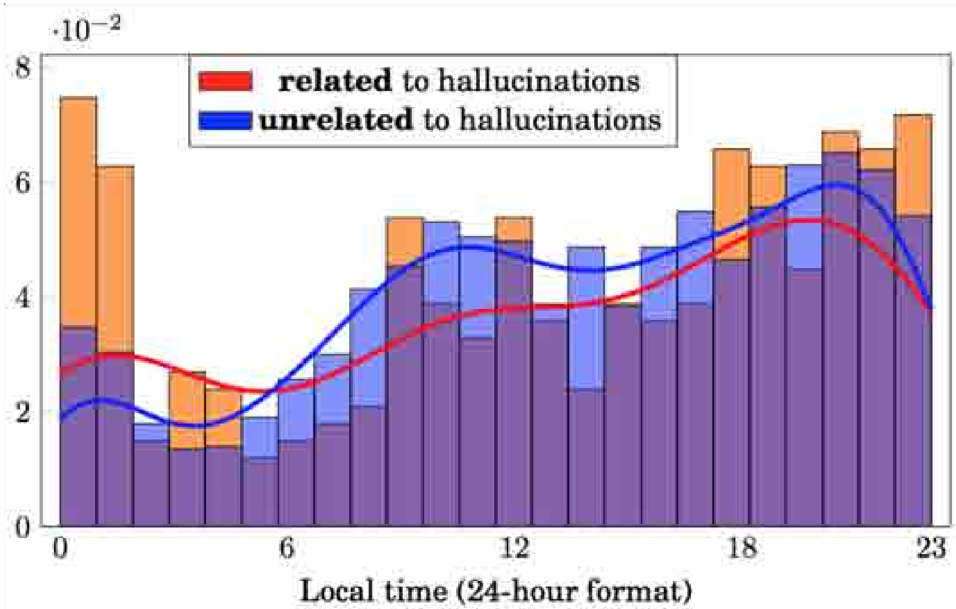
x



Mining healthcare social media

- **Negative sentiments** significantly associated with posts that indicated the occurrence of auditory hallucinations
- Posts linked to auditory hallucinations had a **higher proportionate distribution** between the hours of **11pm** and **5am**





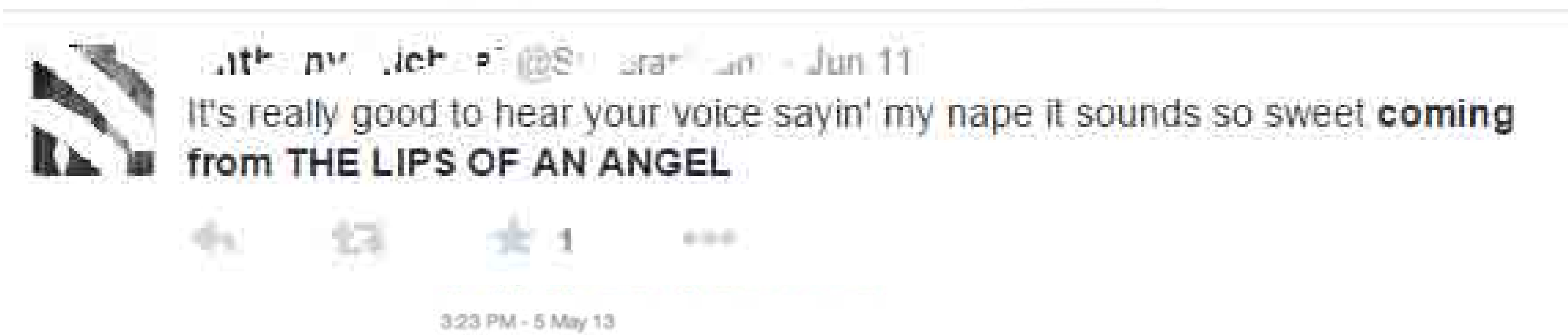
Gaussian kernel density estimate across time

M Belousov et al.: *Mining Auditory Hallucinations from Unsolicited Twitter Posts.* Proceedings of LREC 2016 Workshop.

Mining healthcare social media

- Issues:
  - Layman terminology
  - spelling errors
  - subjective…

- “Interpretation” of patient comments



#WhyWeTweetMH

Understanding why people use Twitter to discuss mental health problems

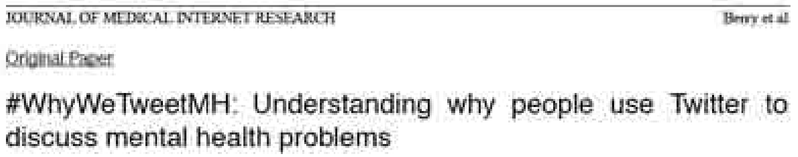
Four main themes:

- Sense of community
  "Because I am with friends even when I am unable to go out."

- Stigma and awareness
  "…begin speaking about what's actually important…"

- Safe space for expression
  "… because I'm never dismissed by my Twitter friends as being over sensitive, needing attention or not making enough of an effort."

- Coping and empowerment
  "My Twitter timeline performs as a sort of mood monitor for myself and those who personally know me…"



JOURNAL OF MEDICAL INTERNET RESEARCH Berry et al

Original Paper

#WhyWeTweetMH: Understanding why people use Twitter to discuss mental health problems

Topics in MH-related posts

Theme	#Posts	Description
BPD	11,880	Forum to discuss aspects of Borderline Personality Disorder either as a sufferer, someone closely related to a sufferer, or someone interested in this disorder
bipolar (*BipolarSOs, BipolarReddit, bipolar*)	41,636	Communities to discuss issues surrounding Bipolar Disorder; while bipolar and BipolarReddit focus on sufferers and their support, BipolarSOs invites contributions from people that are in a relationship with someone suffering from Bipolar Disorder
schizophrenia	4,963	Subreddit to discuss schizophrenia-type disorders and schizophrenia-related issues such as psychosis
Anxiety	57,523	Forum for anything that is related to an anxiety disorder; does not distinguish between sufferer or someone related to a sufferer
depression	197,436	A community for helping anyone struggling with depression; posters are not limited to those who have received a diagnosis by their GP/hospital doctor and the emphasis is on supporting others in their struggle with depression
selfharm (*selfharm, StopSelfHarm*)	17,102	Forums to discuss aspects of people self-harming; while selfharm aims to build a community of sufferers, StopSelfHarm focusses on supporting anyone wanting to stop self-harming even if through a related person

reddit

Characterisation of mental health conditions in social media using Informed Deep Learning

George Gkotsis[1], Anika Oellrich[1], Sumithra Velupillai[1,2], Maria Liakata[3], Tim J. P. Hubbard[4], Richard J. B. Dobson[1,5] & Rina Dutta[1]

Topics in MH-related posts

SuicideWatch	90,518	Forum to support individuals thinking about suicide or people thinking of someone else being at risk of suicide
addiction	4,360	Community to discuss any physical or psychological dependence, e.g. drugs or video games; encourages self post, but does not exclude non-sufferers
cripplingalcoholism	38,241	Community for alcohol-dependent people, with an emphasis on the acceptance of the condition, also stretching to embracing their condition
Opiates (*OpiatesRecovery, opiates*)	65,143	Forums to discuss opiate addiction; while opiates addresses all aspects of the addiction, OpiatesRecovery focusses strongly on supporting everyone wanting to withdraw from opiates; Posting to opiates is restricted to people aged over 18 years
autism	9,470	Forum for anything related to an Autism Spectrum Disorder; provides information and support to anyone facing a diagnosis whether for themselves or someone else
Non-mental health	476,388	Control dataset generated using posts from users on subreddits outlined above, who have posted on other subreddits that are not mental health related

Characterisation of mental health conditions in social media using Informed Deep Learning

George Gkotsis[1], Anika Oellrich[1], Sumithra Velupillai[1,2], Maria Liakata[3], Tim J. P. Hubbard[4], Richard J. B. Dobson[1,5] & Rina Dutta[1]

Text mining patient feedback

- Patient feedback on services
  - treatment effectiveness
  - side effects
  - safety concerns
  - healthcare environment
  - communication and involvement
  - coordination of care

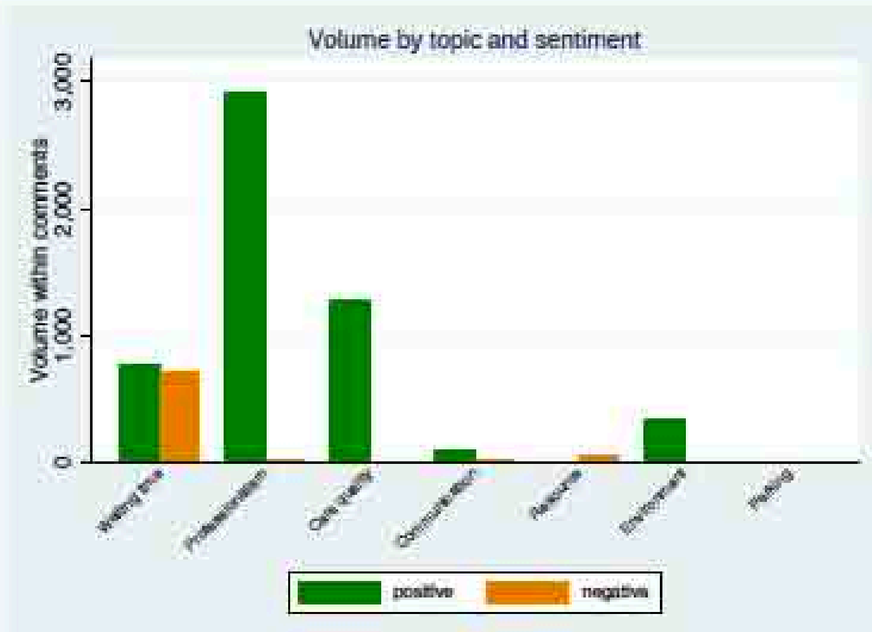
- Use sentiment analysis and topic modelling to identify topics and associated experience

This project is funded by the National Institute for Health Research (NIHR) HS&DR programme, project **14/156/16**. The views and opinions expressed are those of the authors and do not necessarily reflect those of the NIHR, the NHS or the Department of Health

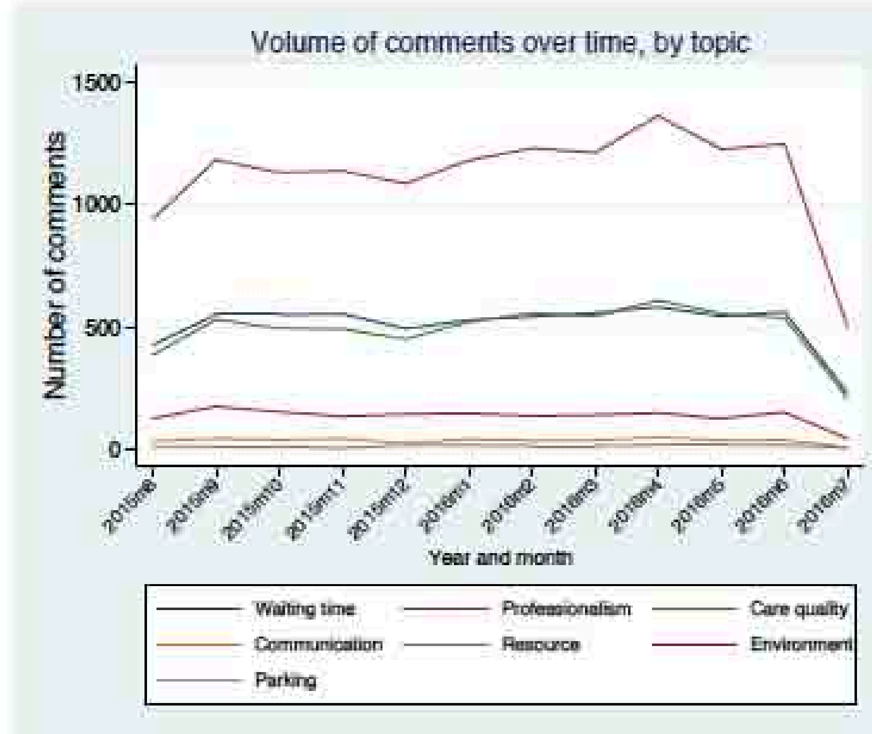
NHS
National Institute for Health Research

Text mining patient feedback

1 Last 3 months, aggregated



2 Last 12 months, time trends



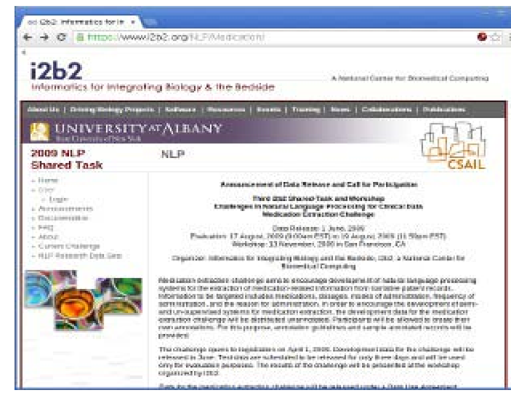
This project is funded by the National Institute for Health Research (NIHR) HS&DR programme, project **14/156/16**. The views and opinions expressed are those of the authors and do not necessarily reflect those of the NIHR, the NHS or the Department of Health

NHS
National Institute for Health Research

State of the art

i2b2

i2b2 challenges in information extraction from clinical narratives



- De-identification (2006, 2014)
- Smoking Status (2006)
- Obesity and disease status (2008)
- Medications (2009)
- Concepts, assertions, relations (2010)
- Coreference resolution (2011)
- Temporal relations (2012)
- Risk factors (2014)
- Symptom severity in psychiatric notes (2016)

State of the art

Progression of CAD Risk Factors in Diabetic Patients Results

	Min	Mean	Median	Max	Std. Dev.
Micro Precision	.455	.808	.852	.913	.119
Micro Recall	.203	.835	.908	.969	.175
Micro F1	**.305**	**.815**	**.872**	**.928**	**.145**
Macro Precision	.455	.800	.849	.914	.121
Macro Recall	.258	.834	.904	.968	.162
Macro F1	.365	.812	.870	.928	.137

i2b2 2014 challenge

State of the art

F measure	Precision	Recall	Method
0.69	0.71	0.67	Rule based pair selection+CRF+SVM
0.69	0.75	0.64	MaxEnt+SVM+rule based
0.68	0.66	0.71	SVM
0.63	0.76	0.54	SVM+rule-based
0.61	0.54	0.72	CRF
0.59	0.65	0.54	MaxEnt/Bayes
0.56	0.57	0.56	Rule based+MaxEnt
0.56	0.48	0.66	SVM
0.55	0.51	0.59	SVM
0.43	0.34	0.59	MaxEnt

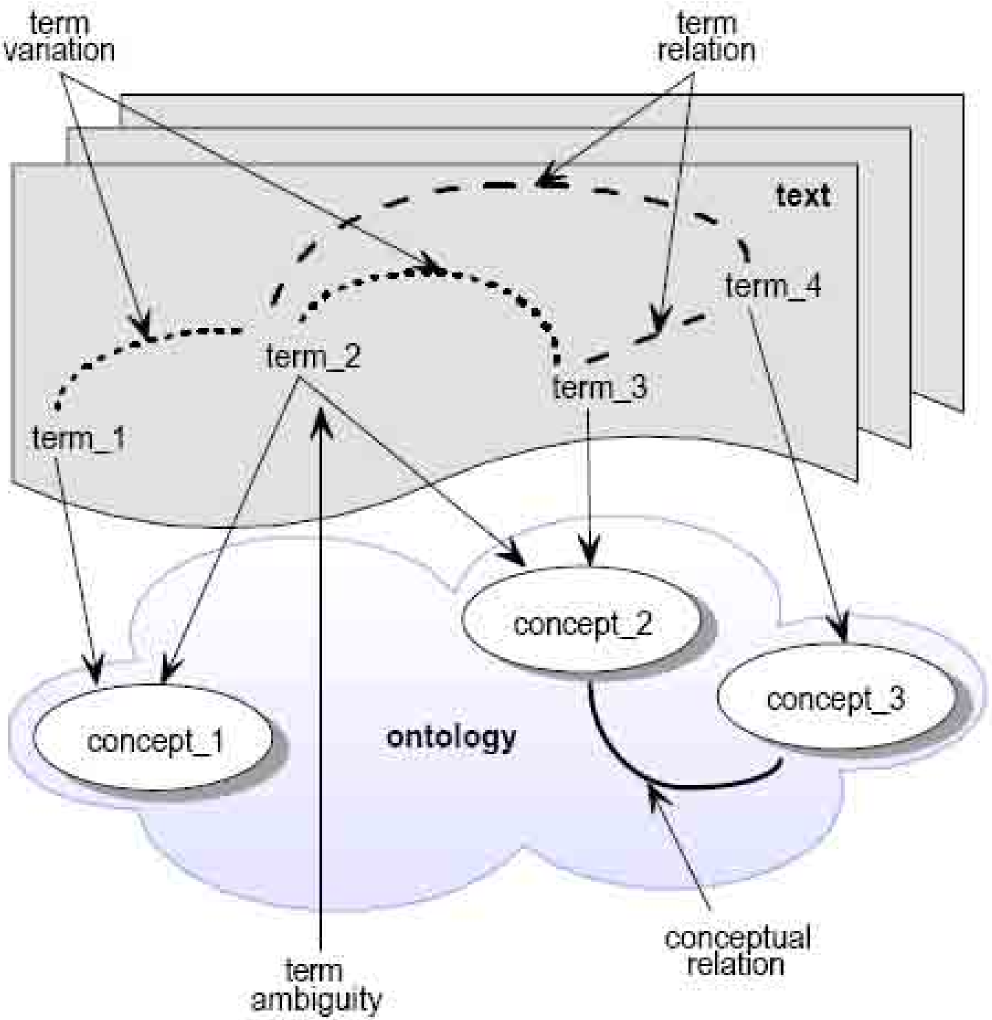
i2b2 2012 – extraction of temporal links

Sun et al. J Am Med Inform Assoc 2013;

Mining clinical narratives

- Challenges
  - Highly condensed text, often without
    - proper spelling
    - proper sentences
    - specific discourse
  - Terminological variability and ambiguity
    - orthographic, acronyms, local conventions
    - mapping to standardised terminologies

Mapping to terminologies



I. Spasic

Automated coding

- E.g. using SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms)
  - must be adopted by all GPs before 1 April **2018**.
  - Secondary Care, Acute Care, Mental Health, Community systems, Dentistry and other systems used in direct patient care by 1 April **2020**.

SNOMED CT
The global language of healthcare

NHS Digital

- Can we automatically code

"Chronic renal impairment (eGFR 44)"

Chronic kidney disease stage 2 (disorder)
SCTID: 431856006

Mining MRI reports

- Convert a radiologist's interpretation of the image into a structured form
- This information is then used by a clinician to support decision making on appropriate treatment

→

HISTORY Injury two weeks ago, ACL and lateral meniscal tear

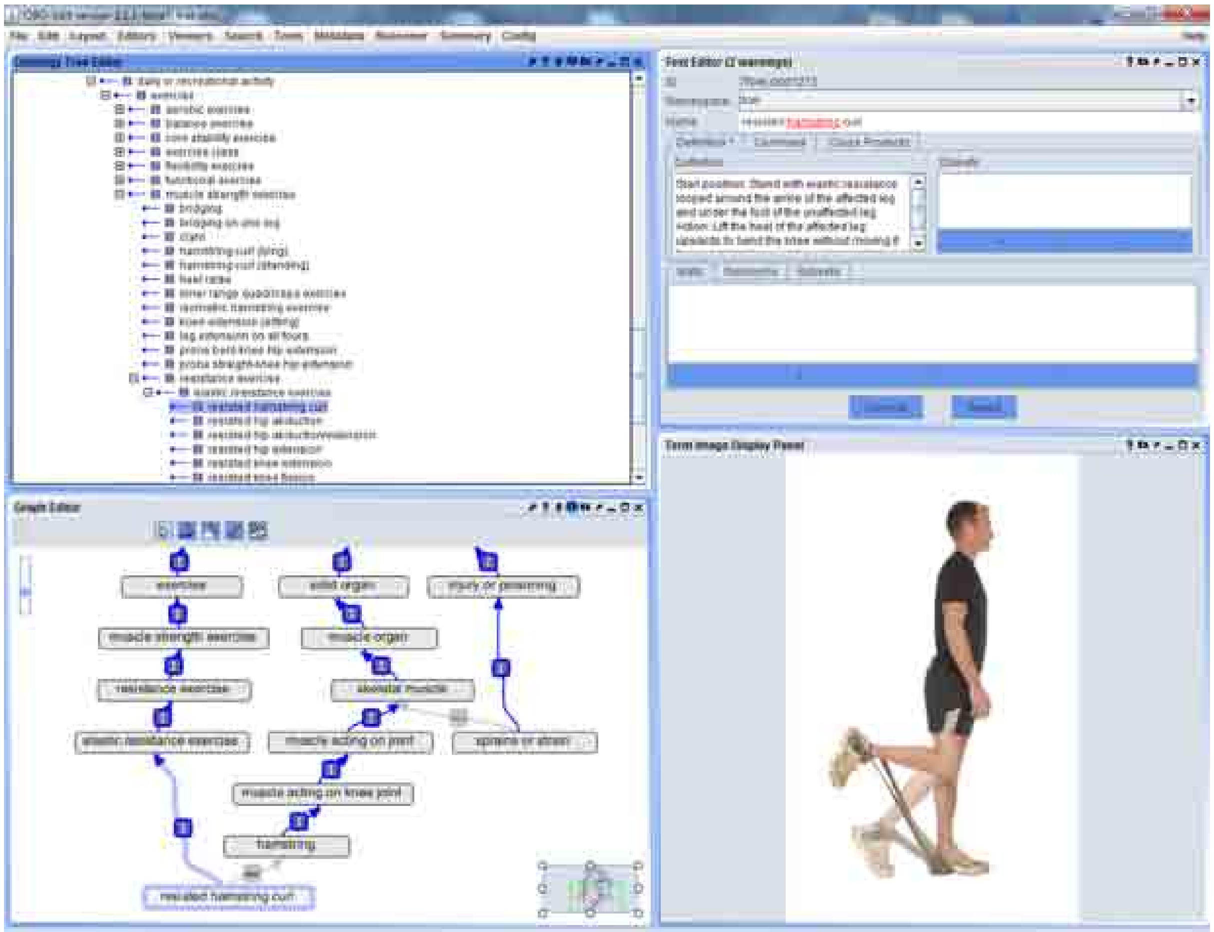
MRI LT KNEE There has been a complete ACL tear in its mid portion. The medial meniscus is intact. There is a radial tear in the lateral meniscus.
The PCL is intact. There is bone marrow oedema in the lateral femoral condyle consistent with trauma from a rotational injury. In addition there is a fragment following cartilage signal intensity lying just medial to the PCL insertion possibly representing a cartilage fragment from the lateral femoral condylar notch. There is a large joint effusion. The lateral ligamentous complex is intact. There is oedema surrounding the MCL consistent with a sprain but the ligament is intact. The posterolateral corner is intact. The patella cartilage is unremarkable.

CONCLUSION Complete ACL tear, radial tear in the lateral meniscus, MCL sprain, depression of the lateral femoral condylar notch with bone marrow oedema and a small cartilaginous fragment at the medial aspect of the PCL insertion.

→

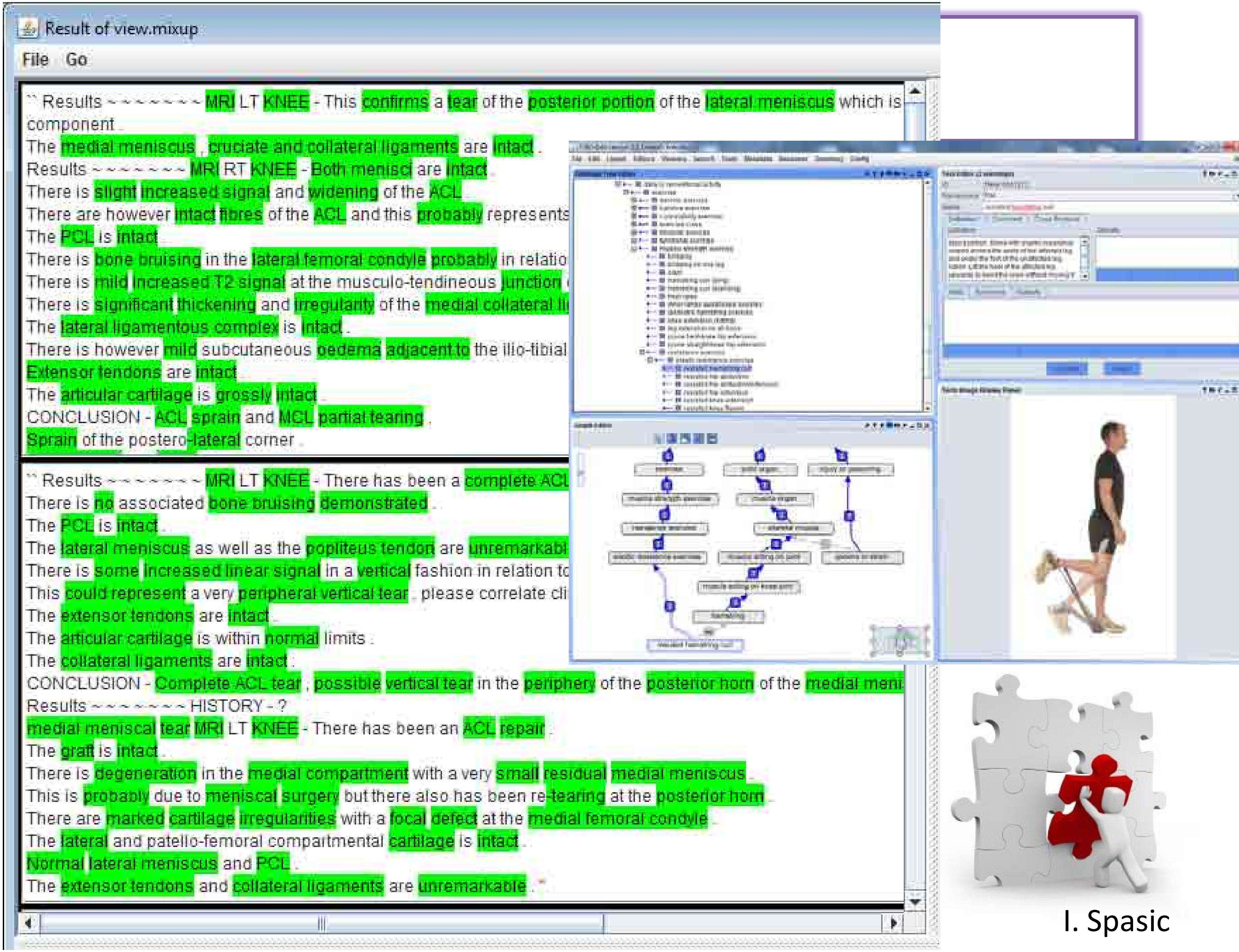
I. Spasic

TRAK ontology



Taxonomy for RehAbilitation of Knee conditions

I. Spasic


Result of view.mixup
File Go
`` Results ~ ~ ~ ~ ~ ~ ~ MRI LT KNEE - This confirms a tear of the posterior portion of the lateral meniscus which is
component .
The medial meniscus , cruciate and collateral ligaments are intact .
Results ~ ~ ~ ~ ~ ~ ~ MRI RT KNEE - Both menisci are intact .
There is slight increased signal and widening of the ACL .
There are however intact fibres of the ACL and this probably represents
The PCL is intact .
There is bone bruising in the lateral femoral condyle probably in
There is mild increased T2 signal at the musculo-tendineous junction
There is significant thickening and irregularity of the medial collateral
The lateral ligamentous complex is intact .
There is however mild subcutaneous oedema adjacent to the ilio-tibial
Extensor tendons are intact .
The articular cartilage is grossly intact .
CONCLUSION - ACL sprain and MCL partial tearing .
Sprain of the postero-lateral corner .
`` Results ~ ~ ~ ~ ~ ~ ~ MRI LT KNEE - There has been a complete ACL
There is no associated bone bruising demonstrated .
The PCL is intact .
The lateral meniscus as well as the popliteus tendon are
There is some increased linear signal in a vertical fashion in relation
This could represent a very peripheral vertical tear , please correlate
The extensor tendons are intact .
The articular cartilage is within normal limits .
The collateral ligaments are intact .
CONCLUSION - Complete ACL tear , possible vertical tear in the periphery of the posterior horn of the medial
Results ~ ~ ~ ~ ~ ~ ~ HISTORY - ?
medial meniscal tear MRI LT KNEE - There has been an ACL repair .
The graft is intact .
There is degeneration in the medial compartment with a very small residual medial meniscus .
This is probably due to meniscal surgery but there also has been re-tearing at the posterior horn .
There are marked cartilage irregularities with a focal defect at the medial femoral condyle .
The lateral and patello-femoral compartmental cartilage is intact .
Normal lateral meniscus and PCL .
The extensor tendons and collateral ligaments are unremarkable . "

I. Spasic

Challenges

- Negation identification
  - there is currently <u>no</u> evidence of a significant meniscal cyst
  - The low signal of the anteromedial bundle seen in a normal ACL is completely <u>absent</u>

- Suspected diagnosis
  - Likely primary Raynaud's in hands and feet

- Family history
  - Her mother's brother was diabetic.

- Patient generated data
  - Once I start moving around or exercise the joint stiffness easies.
  - Constant pain <u>weather</u> sitting or standing.

Challenges

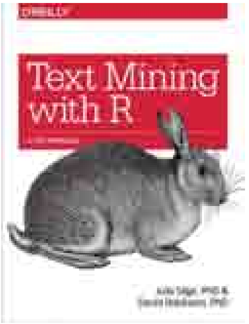
- Coordination		medial and lateral meniscus
- Coreference		the medial meniscus ... the meniscus

- Temporal information extraction
  - recent scan, doesn't feel well recently – is it the same 'recent'
  - check the serum levels in 3 months – 90 days?
  - take 1 tablet with every meal – how many times?

- Approximate expressions
  - pea-sized nodule in the neck – how big is it?

State of the art – systems

- cTAKES - http://ctakes.apache.org/
- Open Health NLP - http://www.ohnlp.org/
- GATE infrastructure
- NLTK
- Text mining with R
- A number of commercial products



State of the art – cTAKES

- cTAKES - http://ctakes.apache.org/

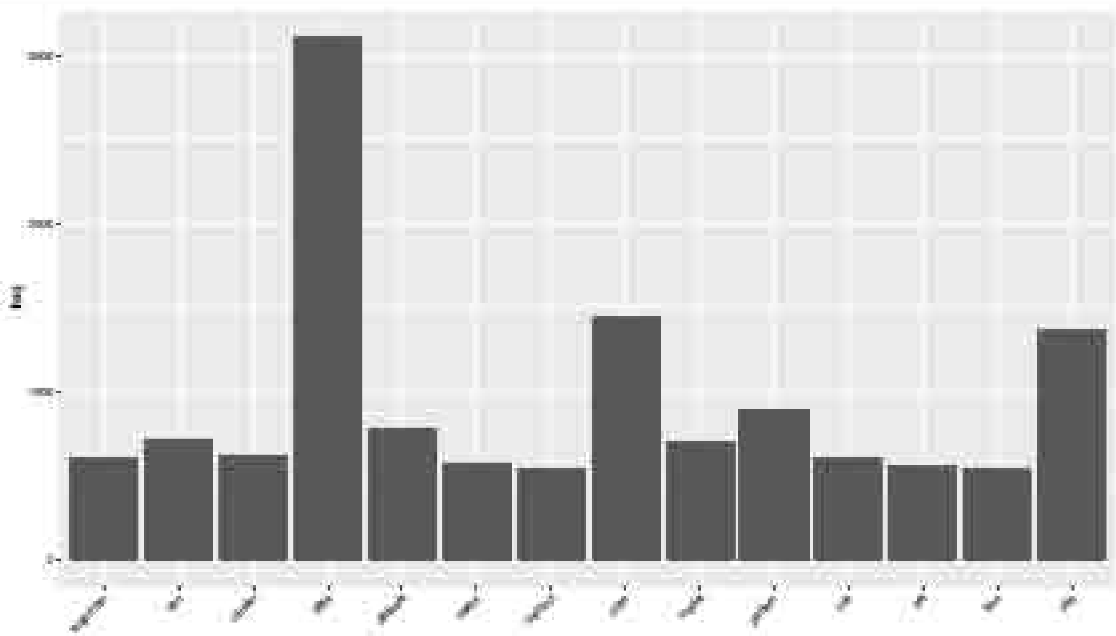
- Demo at

  http://chipweb2.chip.org/cTakes_webservice_trunk/index.html

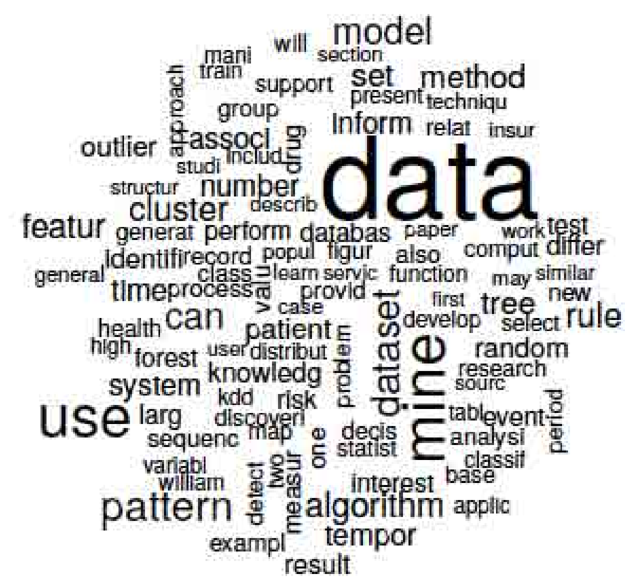
State of the art

- A number of libraries for basic text processing, frequencies, word clouds, finding associations

```
library(ggplot2)
subset(wf, freq>500)                                    %>%
  ggplot(aes(word, freq))                                +
  geom_bar(stat="identity")                              +
  theme(axis.text.x=element_text(angle=45, hjust=1))
```

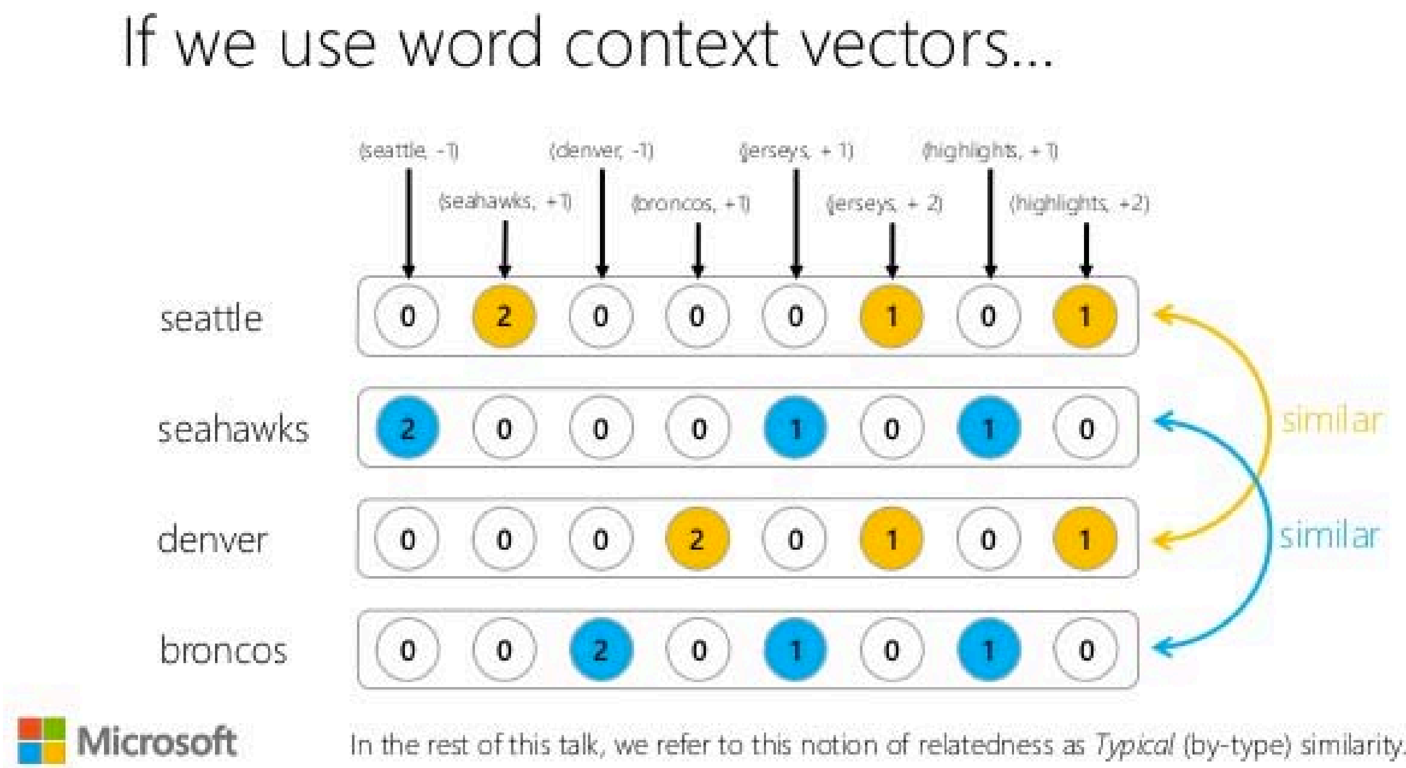


```
set.seed(142)
wordcloud(names(freq), freq, max.words=100)
```



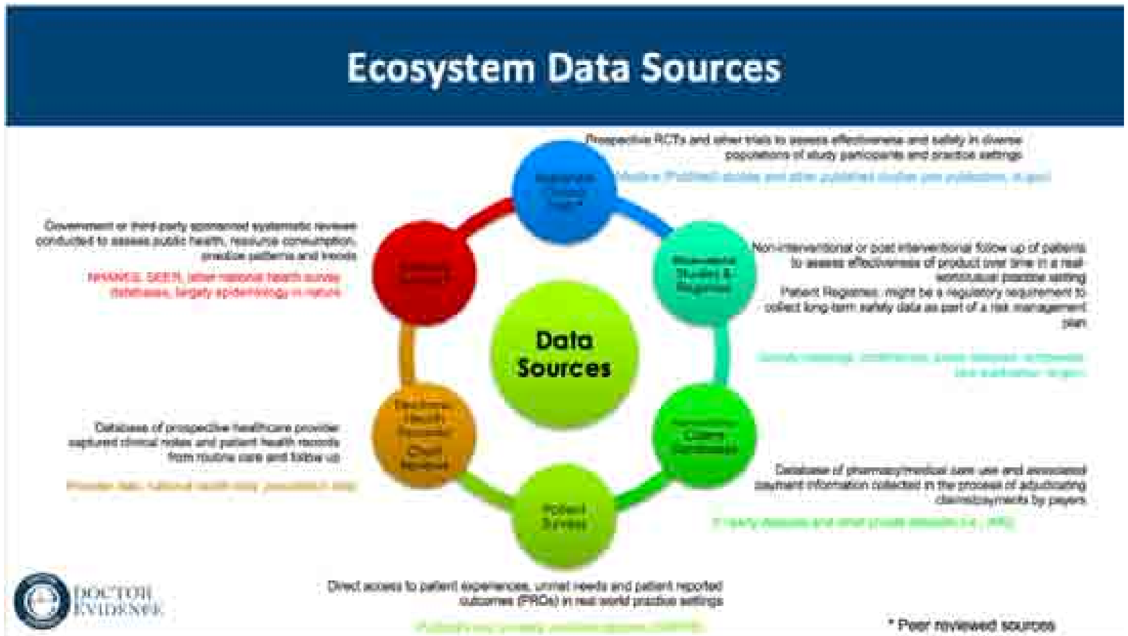
State of the art

- **Word embeddings** – language models where words are represented as vector of numbers, which are learnt from a large corpus



State of the art

- IBM Watson Development Cloud (ex AlchemyAPI)



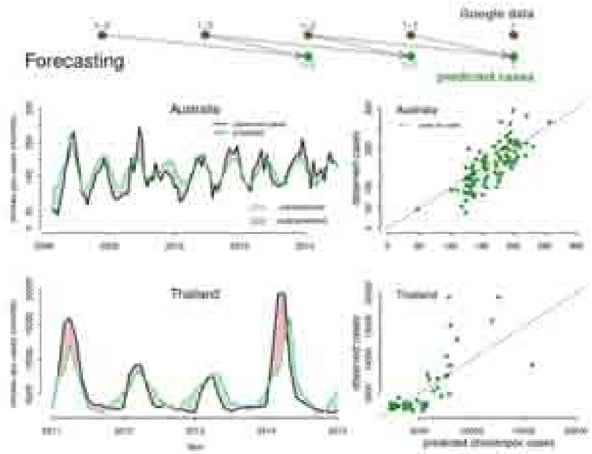
200 medical textbooks and 300 medical journals.

http://drevidence.com/

Support to clinical decision support:

- Parse doctor's query
- Parse EHR
- Parse guidelines, other data to test hypothesis
- Suggest individualised treatment options

Use case scenarios



Health e-Research Centre

Healtex

State of the art

- Some off-the-shelf tools are useful for basic NLP tasks

- But, “there is no such thing as "THE TOOL" for text mining”
  - There will be always the need for tailoring
  - Some tools are more suitable for a given task, but less for the other

- Successful “text mining is 20% engineering, 40% algorithms and 40% science/statistics”
  - there is always going to be new challenges in every new problem you work on (even if it is similar to the previous one)

Group reflection: Needs

- What kind of text mining application would you like to have?
- What would be the opportunities and challenges?

Summary

- Loads of information is in healthcare free text
  - Clinical narrative, social media
  - Guidelines (e.g. NICE)
- Clinical language(s)
  - condensed text, overloaded with terminology
  - spelling errors, abbreviations (local?)
  - implicit information/assumptions
- Healthcare text mining
  - identify (key) entities and relations of interest
  - place the results in **context**

Summary

- What can we do with this data?
  - Support personalised medicine
    - E.g. tailor the therapy for an individual based on social and medical history, environment, allergies, genotype, etc.
  - Improve our understanding of the diseases
    - Identify patterns in genotypes and phenotypes
  - For audit, monitoring and surveillance
    - Addressing some legal obligations

- Text analytics will be an essential part of Learning Health Systems

→ Improve both clinical practice and science

Mental health
Research

Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project

Richard G Jackson[1], Rashmi Patel[1], Nishamali Jayatilleke[1], Anna Kolliakou[1], Michael Ball[1], Genevieve Gorrell[2], Angus Roberts
Richard J Dobson[1], Robert Stewart[1]

Negative symptoms in schizophrenia: a study in a large clinical sample of patients using a novel automated method

Rashmi Patel[1], Nishamali Jayatilleke[2], Matthew Broadbent[3], Chin-Kuo Chang[2], Nadia Foskett[4], Genevieve Gorrell[5], Richard D Hayes[2], Richard Jackson[2], Caroline Johnston[6], Hitesh Shetty[3], Angus Roberts[5], Philip McGuire[1], Robert Stewart[2]

Analysis of diagnoses extracted from electronic health records in a large mental health case register

Yevgeniya Kovalchuk, Robert Stewart, Matthew Broadbent, Tim J. P. Hubbard, Richard J. B. Dobson

European Child & Adolescent Psychiatry
June 2016, Volume 25, Issue 6, pp 649–658

Clinical predictors of antipsychotic use in children and adolescents with autism spectrum disorders: a historical open cohort study using electronic health records

Healtex

- UK healthcare text analytics research network
  - AIM: unlock the evidence contained in healthcare text
- Multi-disciplinary community
  - data/text analysts
  - clinicians, epidemiologists
  - semantic technologies
  - legal and data protection
  - NHS and industry

Healtex

EPSRC
Pioneering research and skills

MANCHESTER 1824 The University of Manchester

KING'S College LONDON

The University Of Sheffield.

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

THE UNIVERSITY OF EDINBURGH

Healtex

- Partners and members

UK Renal Registry – Renal Association

The Farr Institute of Health Informatics Research

University of Liverpool

South London and Maudsley NHS Foundation Trust

HealthUnlocked

Healtex

Arthritis Research UK centre for epidemiology

The Christie NHS Foundation Trust

Big White Wall™

NHS Scotland

NHS Greater Glasgow and Clyde

GIG Cymru NHS Wales – Bwrdd Iechyd Prifysgol Caerdydd a'r Fro – Cardiff and Vale University Health Board

Linguamatics

Manchester 1824 – The University of Manchester

King's College London

The University Of Sheffield.

Cardiff University – Prifysgol Caerdydd

The University of Edinburgh

Healtex

- Workshops
- Datathons
- Hackathons
- Working groups

Mining radiology reports
De-identification of clinical narrative
Mining medication information from free text
Automated coding of clinical narrative
Processing patient generated data
Temporal clinical information extraction
Mining veterinary clinical narrative (Vetext)

- Call for pilot projects and feasibility studies (May 2017 and 2018)

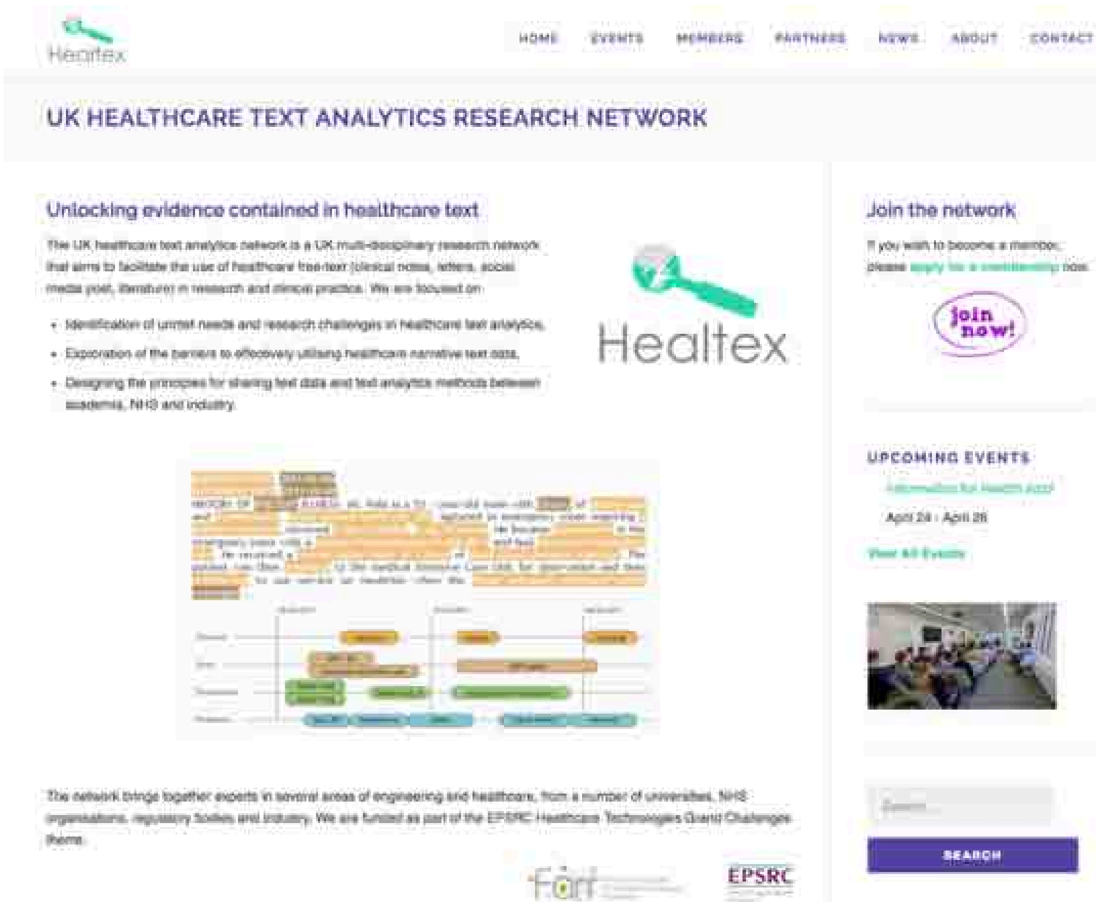
Future workshops

Workshop	Date (TBC)
Processing patient generated data – opportunities, barriers and challenges	October 2017
What's in a clinical narrative – how clinicians compose a free text narrative and why	November 2017
Preserving privacy and facilitating sharing of healthcare free text – is there a best practice for accessing clinical text?	February 2018
Patients' view on using healthcare narrative for research – are patients concerned with sharing the narrative?	April 2018
Challenges of knowledge- and data-intensive text analytics etc. – what are the open problems in processing healthcare text analytics?	September 2018
Integration of clinical text into actionable healthcare analytics – how to make sense of free text?	January 2019

- First UK conference on healthcare text analytics – early 2018
- Workshop at Informatics for Health conference: "Extracting evidence from clinical free text: opportunities and challenges"

  Tuesday 9:30 (Exchange 2)

Join the network



www.healtex.org