Clinical Natural Language Processing: Unlocking Patient Records for Research

Mark Dredze

Computer Science

Malone Center for Engineering Healthcare
Center for Language and Speech Processing

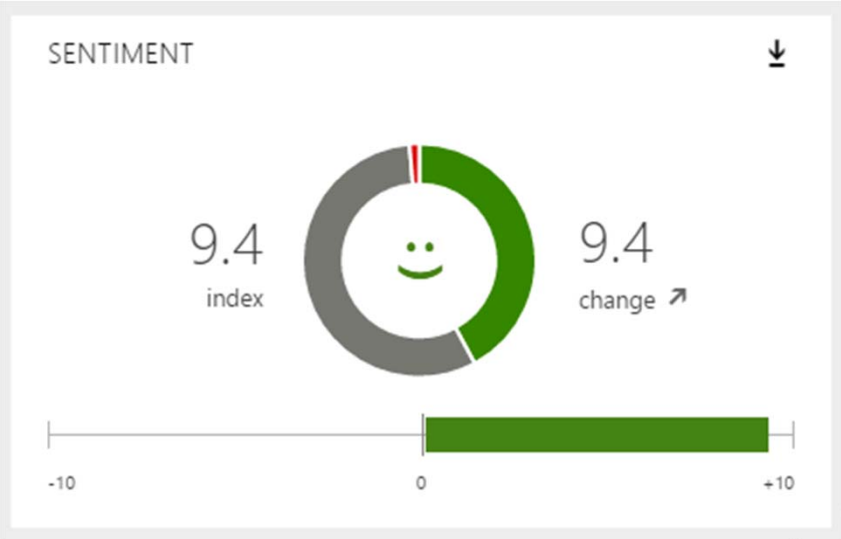
Natural Language Processing

- The computer science discipline that studies language and computers
- Computational Linguistics
  - The study of language aided by computers
- Human Language Technology
  - The development of new technology (algorithms, software, resources) that automate the processing of language

NLP in Industry

Google Translate

Microsoft Cognitive Services

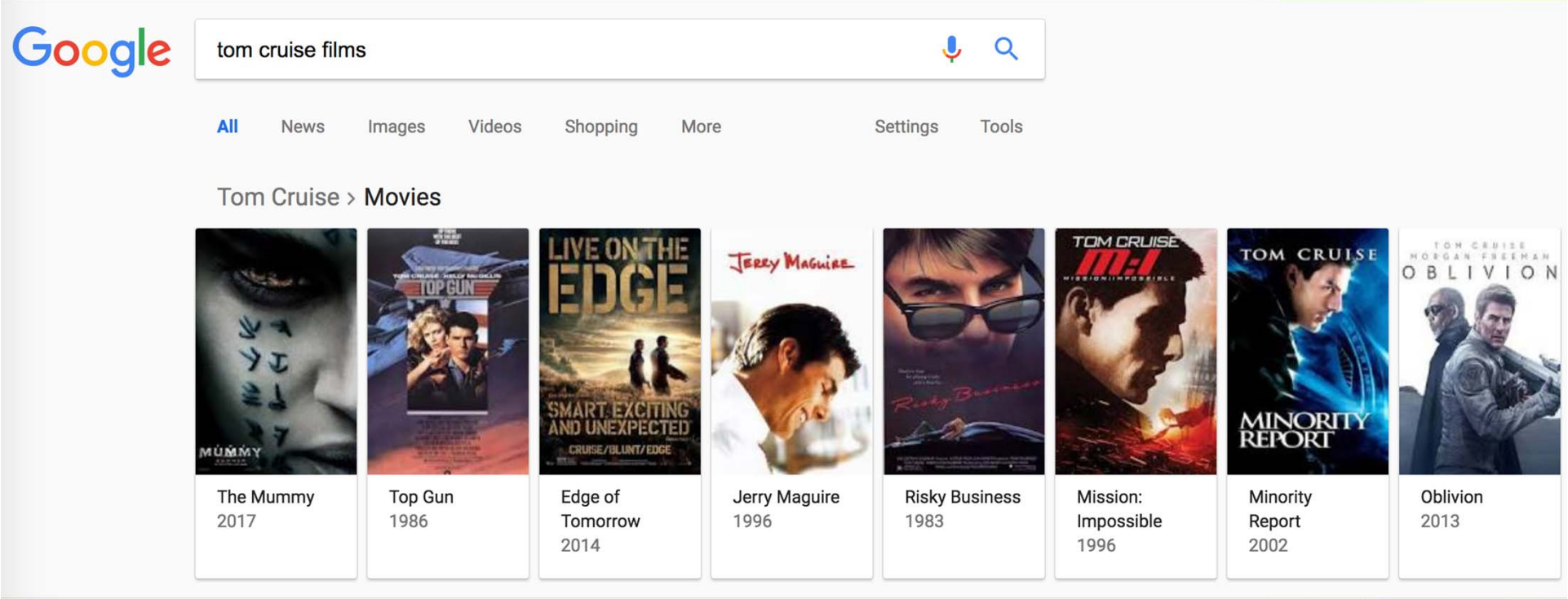


Examples of NLP

- Question Answering
  - What is the methyl donor of DNA (cytosine-5)-methyltransferases?
  - Where in the cell do we find the protein Cep135?
  - Is rheumatoid arthritis more common in men or women?

Examples of NLP

- Information Extraction



https://jamanetwork.com/journals/jamaoncology/fullarticle/2517402

Examples of NLP

- Information Extraction

Section identification
Separates report into "chunks" with a section category

Coreference resolution
Determining that "Mr. Xxxx," "he," and "his" refer to the same person is a coreference task

History of present illness

History of present illness
Mr Xxxxx is a YY-year-old male referred to us by Dr Xxx for evaluation of a new central liver mass found on surveillance imaging for hepatitis B.
He has been followed with yearly ultrasonography of the abdomen and his most recent ultrasonography on DD/MM/YYYY revealed a 7.2-cm mass in the medial right lobe without evidence of ductal dilation.
This was further characterized with multiphase CT on the same day and lesion revealed imaging characteristics consistent with HCC.

Allergies

Allergies
NO KNOWN DRUG ALLERGIES

Medication

Medications
Lisinopril, 60 mg daily
Ranitidine, 150-mg BID

Medical history

Medical history:
Cardiovascular: HTN, valvular disease, tricuspid and mitral valve regurgitation with preserved function
Endocrine: DM
Past liver disease: Hepatitis B
Hepatitis risk factors: None

Surgical history

Surgical history
None

Family history

Family history
Mother: HBV, lung cancer
Father: HTN
Brother: Melanoma

Temporal extraction
Identifying and relating temporal expressions such as "YY year," "DD/MM/YYYY," and "same day"

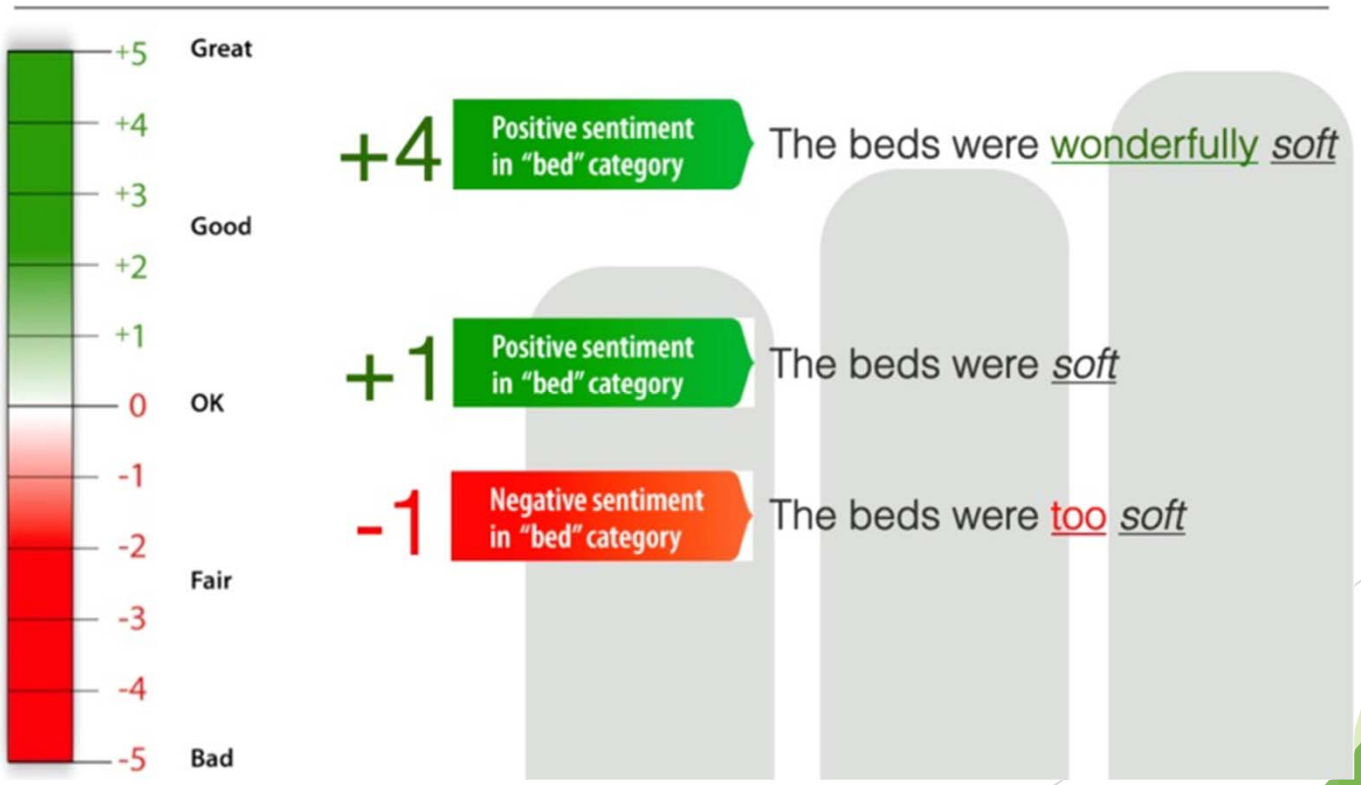
Medication information extraction
Drug: Lisinopril
Strength: 60 mg
Frequency: daily
Drug: Ranitidine
Strength: 150 mg
Frequency: BID

Family history extraction
Family member: Mother
Finding: HBV
Finding: Lung cancer
Family member: Father
Finding: HTN
Family member: Brother
Finding: Melanoma

https://jamanetwork.com/journals/jamaoncology/fullarticle/2517402

Examples of NLP

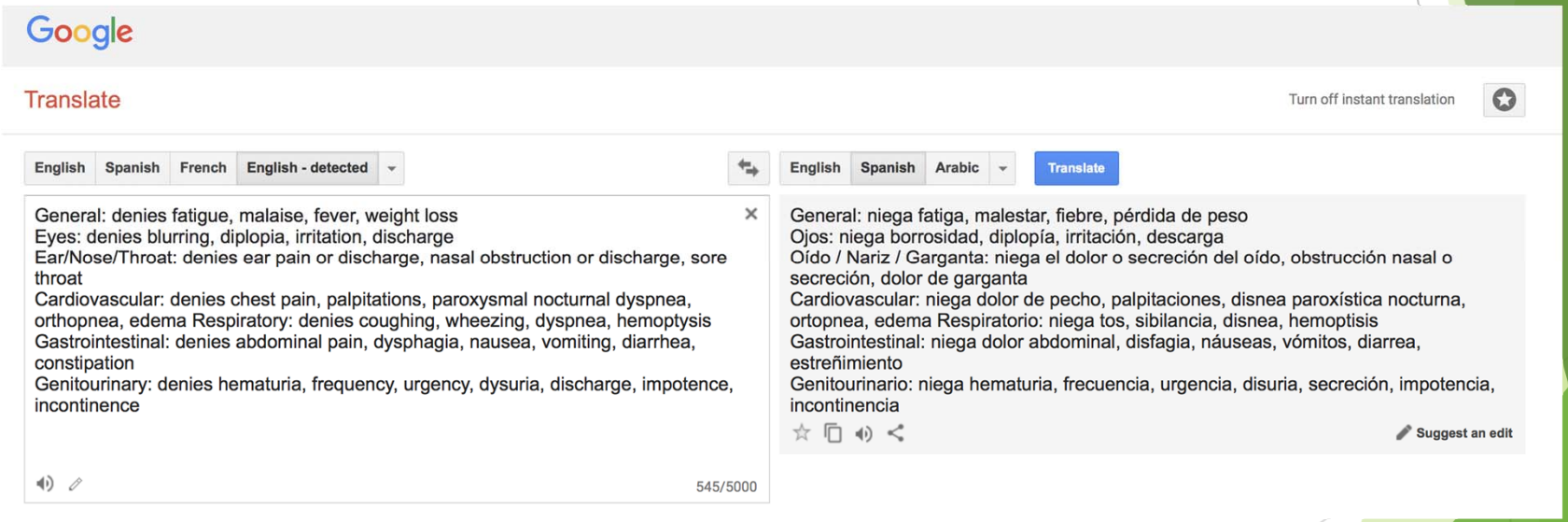
- Sentiment Analysis



https://www.revinate.com/blog/2012/09/understanding-voice-of-the-customer-data/

Examples of NLP

- Machine Translation



https://www.ahrq.gov/professionals/prevention-chronic-care/improve/system/pfhandbook/mod8appbsteveapple.html

The Statistical Revolution

- Before 1990
  - NLP systems are rule based
  - Knowledge engineering
- Starting in 1990s
  - We suddenly get lots of actual data
  - Focus on statistical models, and estimate parameters on data
- Deep Learning
  - Statistical methods with millions of parameters estaimted from data
- Key: Training data!
  - Language data is everywhere

https://http://bigdata.rutgers.edu/news/significant-benefits-of-big-data-analytics-in-healthcare-industry

The Statistical Revolution

- Statistical revolution hitting clinical data starting in 2010

https://http://bigdata.rutgers.edu/news/significant-benefits-of-big-data-analytics-in-healthcare-industry

Where Does Language Appear In Medicine?

- Clinical notes (from physicians, labs, radiology, …)
- Patient diaries
- Messages among doctors or between doctors & patients.
- Medical literature
- Spoken doctor patient interactions
- …

What Can NLP Do?

- Information organization
  - Sort information by topic, etc.
  - High level views of data
  - Identifying relations between entities across dataset
  - Model correlations between text and structured fields
- Information Extraction
  - Extract entities, relations, events, outcomes
  - Produce structured knowledge from text
  - Reasoning from text
  - Link entity mentions across documents to each other and KB
- Information Access
  - Language translation, speech transcription

Uses of Clinical NLP

- Supporting research
  - Tools and methods that enable support of research using NLP
  - Extracting or structuring language data for use in research
- Improving Care
  - Important in clinical decision support systems

Clinical NLP Tasks

- Basic note processing
  - Segmentation, syntax, text normalization, processing abbreviations, temporal expressions, numerical values
- Entities
  - Entity extraction: identify names of important entities in text
- Concepts
  - Concept linking: connect mentions of concepts to ontologies
  - Phenotyping
- Beyond
  - Summarization
  - …

General clinical NLP

- De-identification of clinical notes
- Medication intake information (esp. over-the-counter)
- Temporal information (e.g. dates, duration)
- Numerical values of specific variables (e.g. labs, vitals)
- Suspicious breast cancer lesions
- Detection of smoking status

Center for Language and Speech Processing

- World leader in NLP
- Understand how human language is used to communicate ideas/thoughts/information.
- Develop technology for machine analysis, translation, and transformation of multilingual speech and text.
- ~13 primary faculty, 10 secondary, 60 graduate students, 6 postdocs

Malone Center for Engineering in Healthcare

- Established in 2016 to promote the user of engineering methods to improve healthcare
- Accelerate development of research-based innovations in healthcare
- 29 affiliated faculty

Center for Clinical Natural Language Processing (C2NLP)

- Founded March 2018
- iCore (ICTR) center focused on NLP innovation and tool development
- Sister center to Center for Clinical Data Analysis (CCDA)
  - Delivery of data as a service
- C2NLP Goals:
  - Enable CCDA to provide NLP data as a service
  - Clinical NLP research as a service
- Collaboration with the JHUAPL Precision Medicine Analytics Platform

C2NLP Goals

- Data access
- Tools
- Best practice
- Community for cNLP research at JHU
- Public face of this research area
- Bring together Whiting, Medicine, Bloomberg, APL

Motivation: Requests for NLP to CCDA

- Information Extraction
  - Find me all records that record a result of test X with value Y
- NLP Tool Evaluation
  - Which is the right tool for our work?
- General (i.e. non-clinical) NLP
  - Can you help us analyze this language dataset?

Information Extraction

- **Disease Based Cohort**
  - Cohort to examine risk factors for end organ disease
  - Identify history of conditions and risk factors reported in clinical text
- **Test Results Based Cohort**
  - Correlation between quantitative scores for medical test and diagnostic exams.
- **Rare Disease Mentions**
  - No ICD code to indicate many rare, or not well defined, diseases and conditions
  - Conditions mentioned in clinical notes in many different ways

NLP Tool Evaluation

- Performance of **cTAKES** (clinical Text Analysis and Knowledge Extraction System)
  - Entity Recognition on different types of cancer pathology reports
- Evaluating NLP tools on MRI, pathology, and clinic reports.

General NLP Applications

- Analysis of the variations in language use by doctors
  - How do doctors talk about different types patients
  - How do different doctors talk about the same topics
  - Content analysis of types of language use
- Measure the effects of different disorders on language use
  - Consider samples of language data collected from patients
  - How does language vary over time for patients receiving certain treatments, or who have received specific diagnoses
  - Lexical richness, syntactic complexity, readability scores

C2NLP services

- General clinical NLP
  - Innovation to support CCDA
- Research as service
  - Need NLP experts for a research project or proposal? We have them!
- Large-scale clinical notes processing
  - What can we learn by considering millions of records at scale

Delivery methods

- HIPAA-compliant servers (mainly PMAP)
- Project-specific environment (Docker containers)
- Depending on the request, any combination of:
  - Raw or processed data
  - NLP packages
  - Data analysis tools

Come Talk to Us

- Founding: March 2018

- Director: Mark Dredze
- mdredze@cs.jhu.edu
- http://www.dredze.com