Text Analysis and Retrieval

1. Introduction

Prof. Jan Šnajder

University of Zagreb
Faculty of Electrical Engineering and Computing (FER)

Academic Year 2021/2022

Creative Commons Attribution–NonCommercial–NoDerivs 3.0

v2.0

Teaching staff

Lecturers:

- Prof. Jan Šnajder

TAs:

- Josip Jukić

Student TAs:

- Lucija Arambašić, Sara Bakić, Rino Čala, Marko Čuljak, Antonijo Marijić, Filip Wolf

Course web site:
`http://www.fer.hr/predmet/apt`

Powered by...

TakeLab

Text Analysis and Knowledge Engineering Lab

`http://takelab.fer.hr`

TL
TakeLab

After today's lecture, you'll...

1. Know what this course is about and be glad that you've enrolled it
2. Know what are the topics that we want to cover
3. Hopefully become intrigued by some of them
4. Know what you need to do to earn your credits

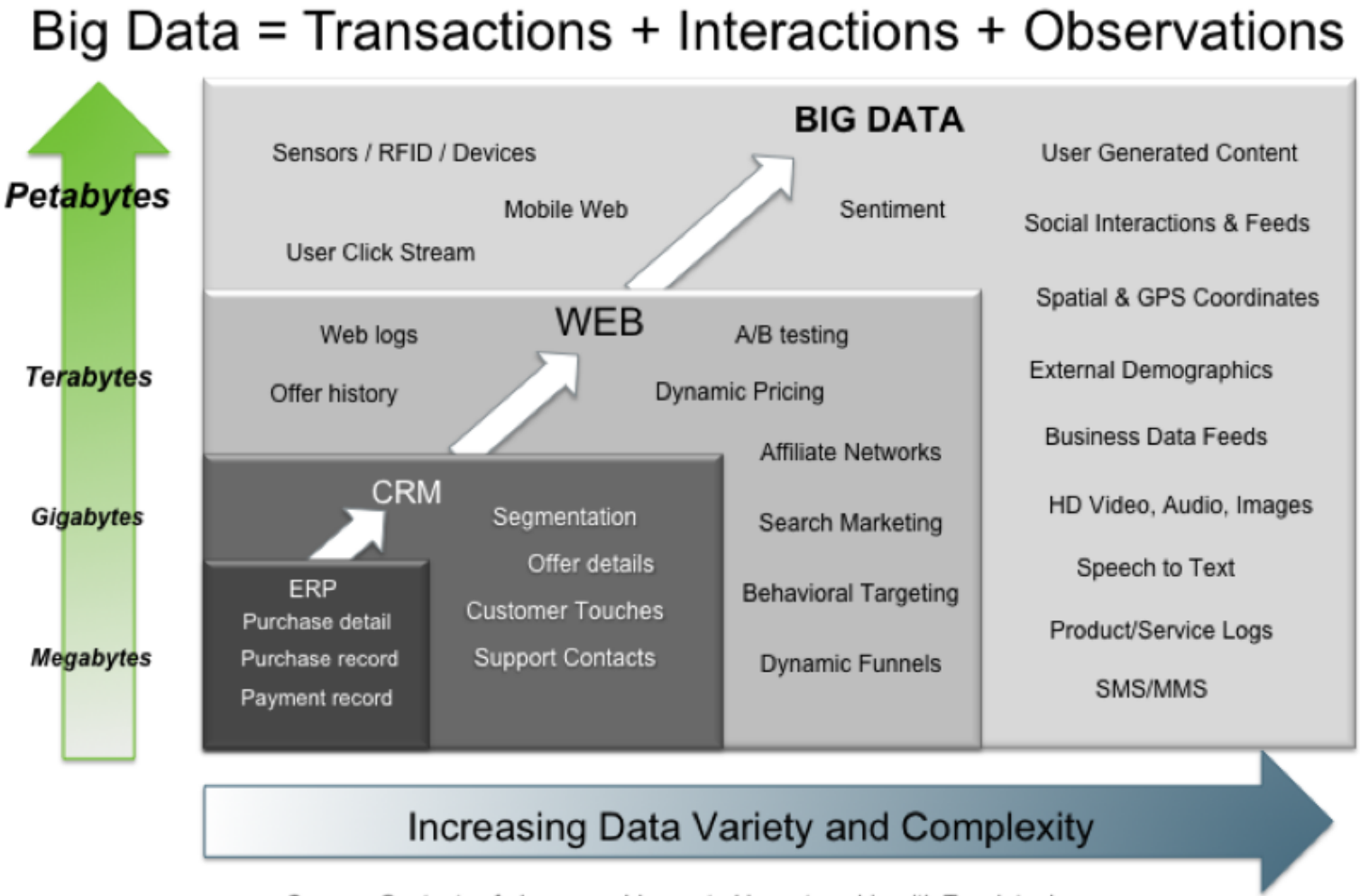
Outline

1. Why this course?
2. Topics
3. Organization

What is this course about?

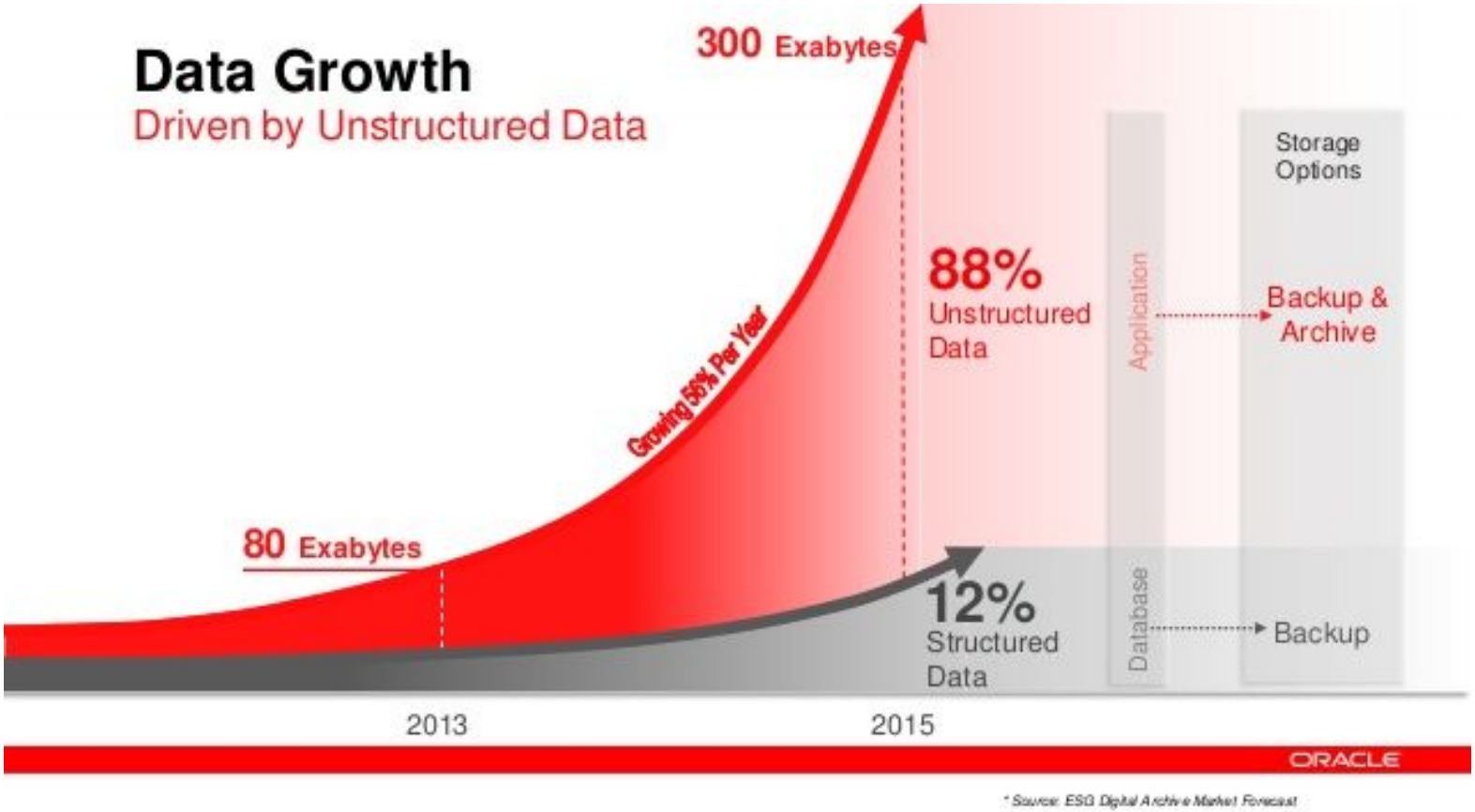
- Text is everywhere (books, news, emails, tweets, reviews,. . .)
- Most human knowledge is stored in unstructured, textual form
- The amount of text data is vast and rapidly growing
- We need systems that address diverse information needs of the users and extract information from large volumes of unstructured data
- Text analysis and retrieval systems are an indispensable part of modern ICT infrastructure

Big Data

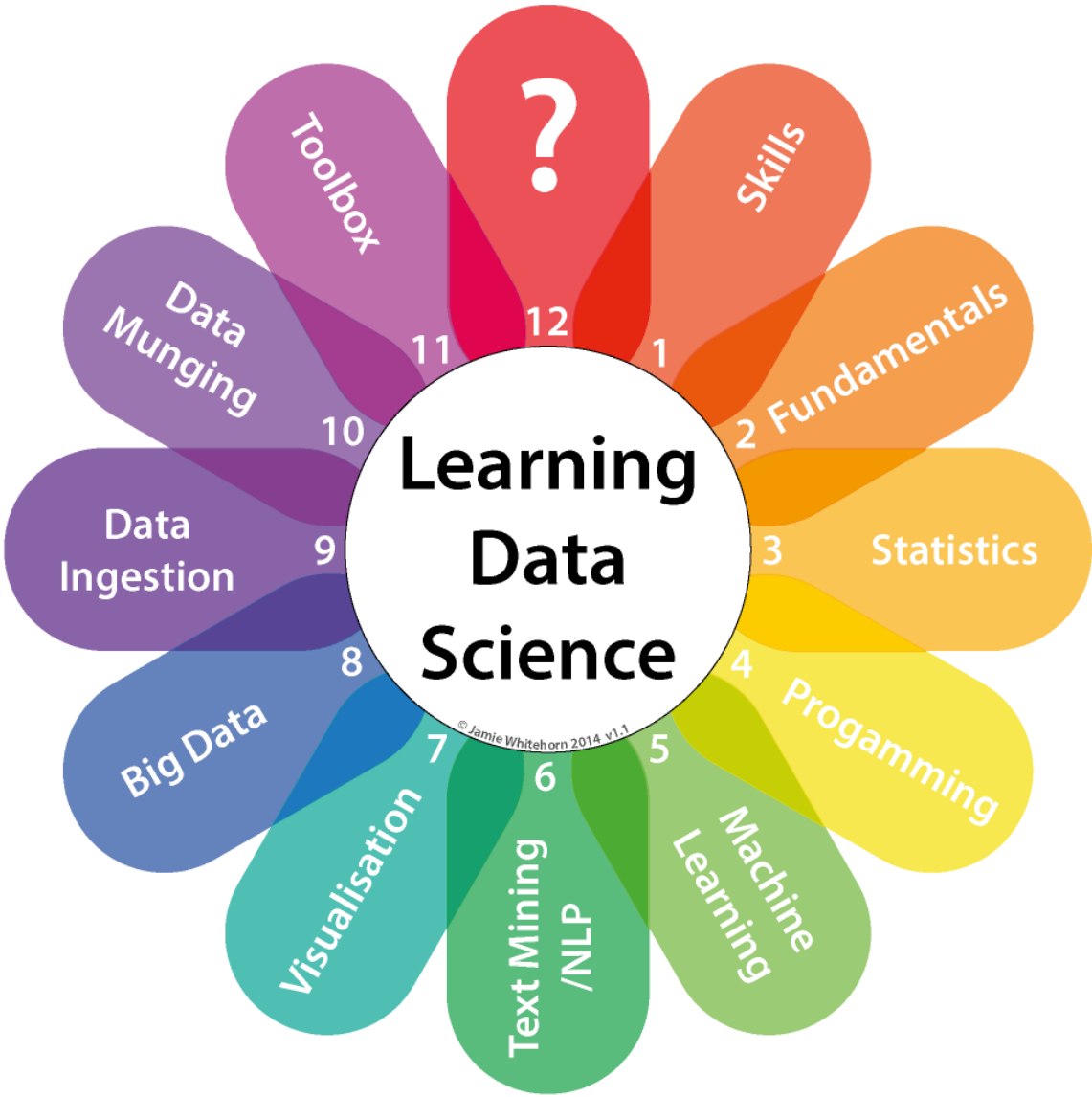


***Source**: Contents of above graphic created in partnership with Teradata, Inc.*

Unstructured data

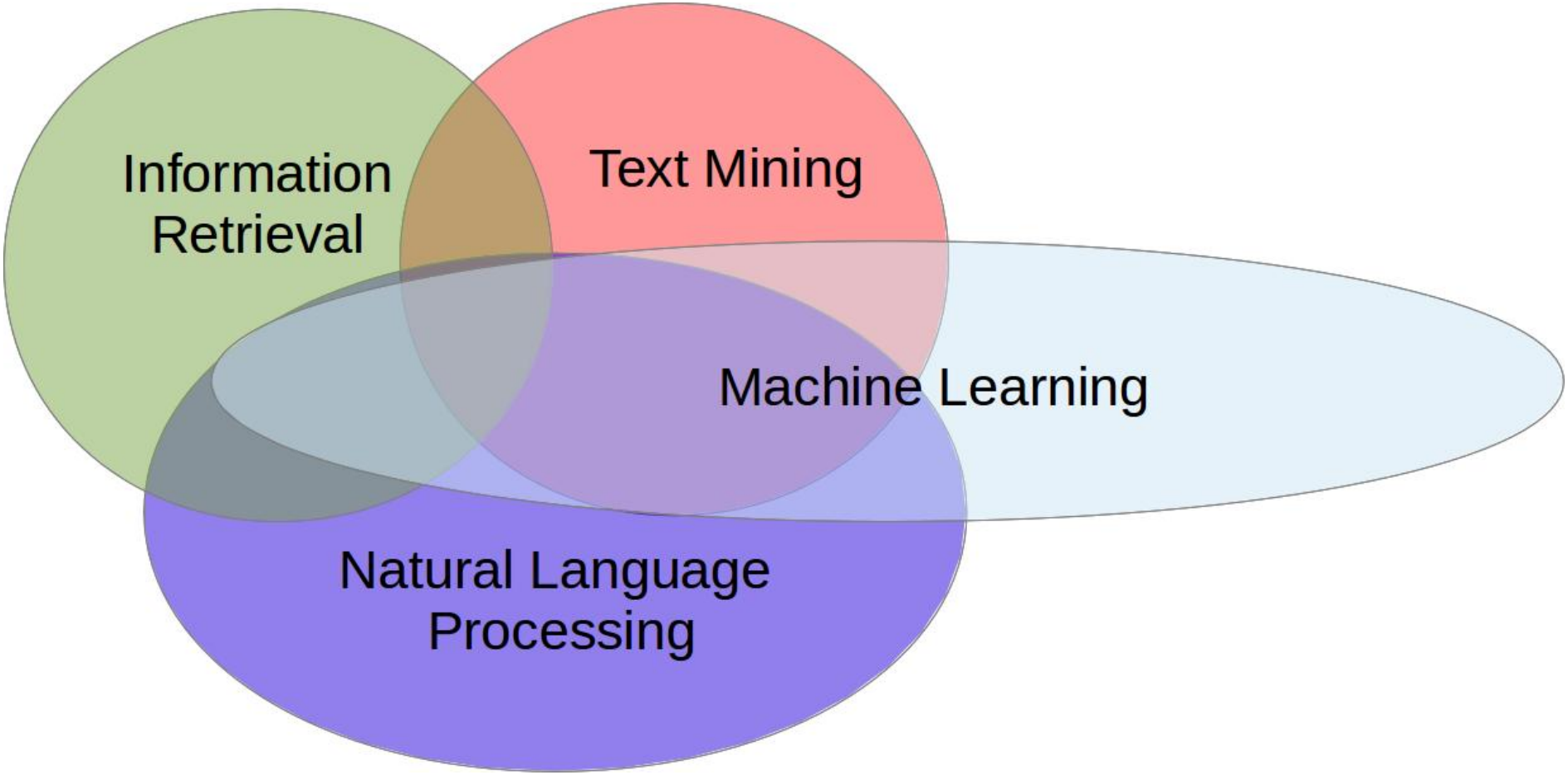


Data Science



http://www.exploringdatascience.com/the-data-science-clock/

What's involved?



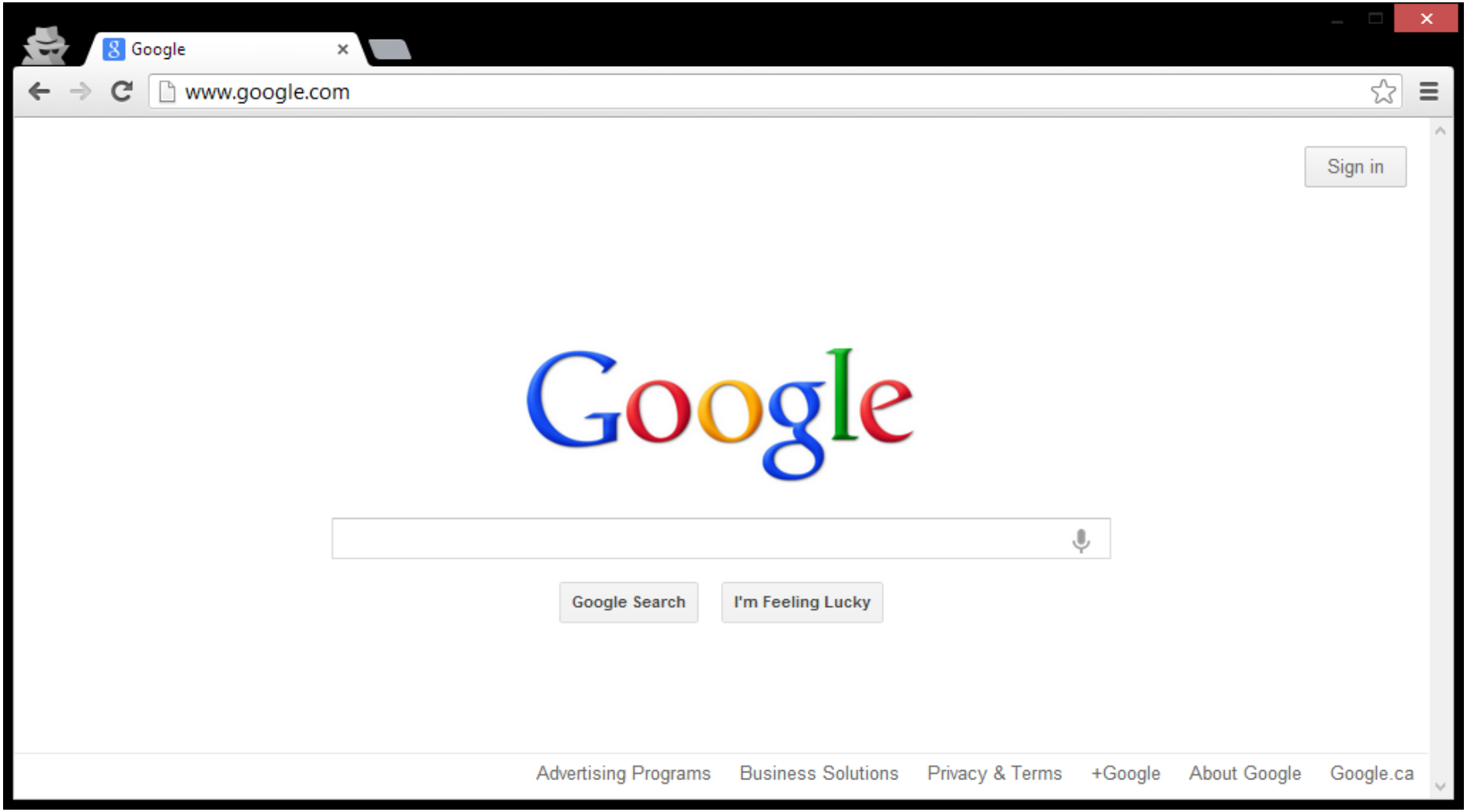
Information retrieval

Information Retrieval (Manning & Schütze 2008)

Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers).

- Web search
- Personal search (emails, etc.)
- Enterprise search
- Domain-specific search

Web search


Google
www.google.com
Sign in
Google
Google Search
I'm Feeling Lucky
Advertising Programs
Business Solutions
Privacy & Terms
+Google
About Google
Google.ca

Domain-specific search

ClassifiedAds.com

Log in Help Post an ad (free)

All Categories | small parrot cage | in Earth | SEARCH

Bird Cage Suppliers Accepting Orders in **Small** Volume
small-order.hktdc.com/ Reduce risk in sourcing! Place order in **small** quantity @ hktdc.com

Sponsored link related to: **small parrot cage**

Used Bird and Parrot Cages, good condition
PLS. EMAIL FOR APPT. TO VIEW ALL! Cleaning out my garage of used bird cages ($5-40), iron parrot cages

large parrot cage, 40"wide, 32"deep, 74" high, almost new
green large parrot cage, just bought, too big for house

parrot cage
Parrot cage. Good condition Comes with two bowel feeders and wooden bench.

Parrot Cage With toys and Dishes
Parrot cage with toys and dishes. If interested please text me and I will send you a picture. Or just call me to

NO IMAGE

Parrot Cage
Black Wrought iron Parrot cage, about 5 ft. tall, fits in corner. Good condition. 561-389-5347

NO IMAGE

Large Parrot Cage For Sale
LARGE WHITE 5 FOOT PARROT CAGE EXCELLENT CONDITION TOP PERCH WITH LOWER SHELVE FOR STORAGE

LARGE Parrot cage for sale
Large Parrot Cage for sale with top perch and lower storage area comes with ladder for bird to climb back into

wanted parrot cage 3 feet wide black willing to pay 150$
Wanted Parrot cage willing to pay 150 Balck 3 feet wide

NO IMAGE

Small Plastic cage with Hermit crab
Purchased this summer by my oldest daughter for near $35, (she is onto other activities.) Has large hermit

PARROT CAGE, TOYS, BOOKS, ETC.
ALL YOU NEED NOW IS THE PARROTnbsp; This Dometop cage provides your pet with a new home that is both

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 ··· >

Create your own ad in Earth . It's easy and free!

Follow, Like or add us to your Google + circles

Tweet 0 Follow

Like 0

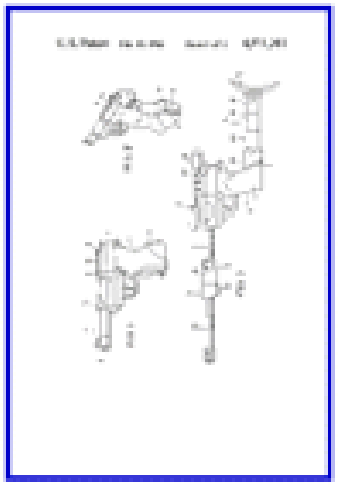
g+1 0 Follow 1.2k

Domain-specific search

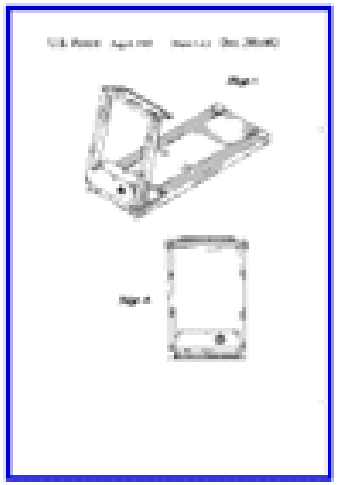
Google™ Patent Search BETA

Search Patents | Advanced Patent Search | Google Patent Search Help

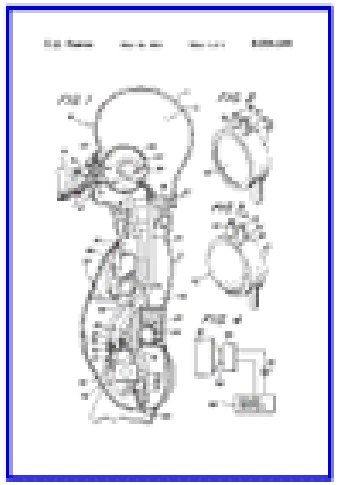
Search over 7 million patents.



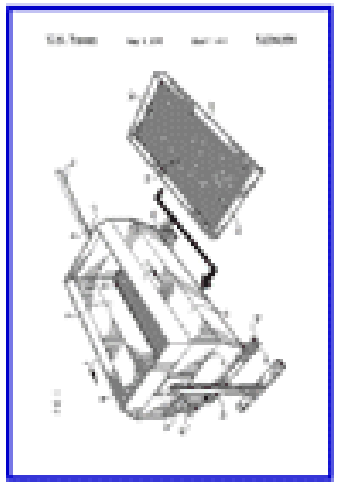
Toy gun convertible into robotic-humanoid form



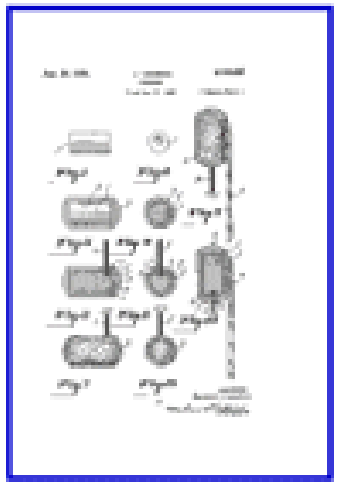
Treadmill



Doll having delayed wetting and crying action



Chicken breading machine



FIREWORKS

Information retrieval problem

- User has an information need, a topic about which she desires to know more
- The user submits a query in an attempt to communicate the information need
- The system compares the query against the document index and retrieves the documents that match the query
- A document is relevant if it is one that the user perceives as containing information of value with respect to their personal information need
- Ideally, the system retrieves all relevant documents and no irrelevant documents

Information retrieval problem

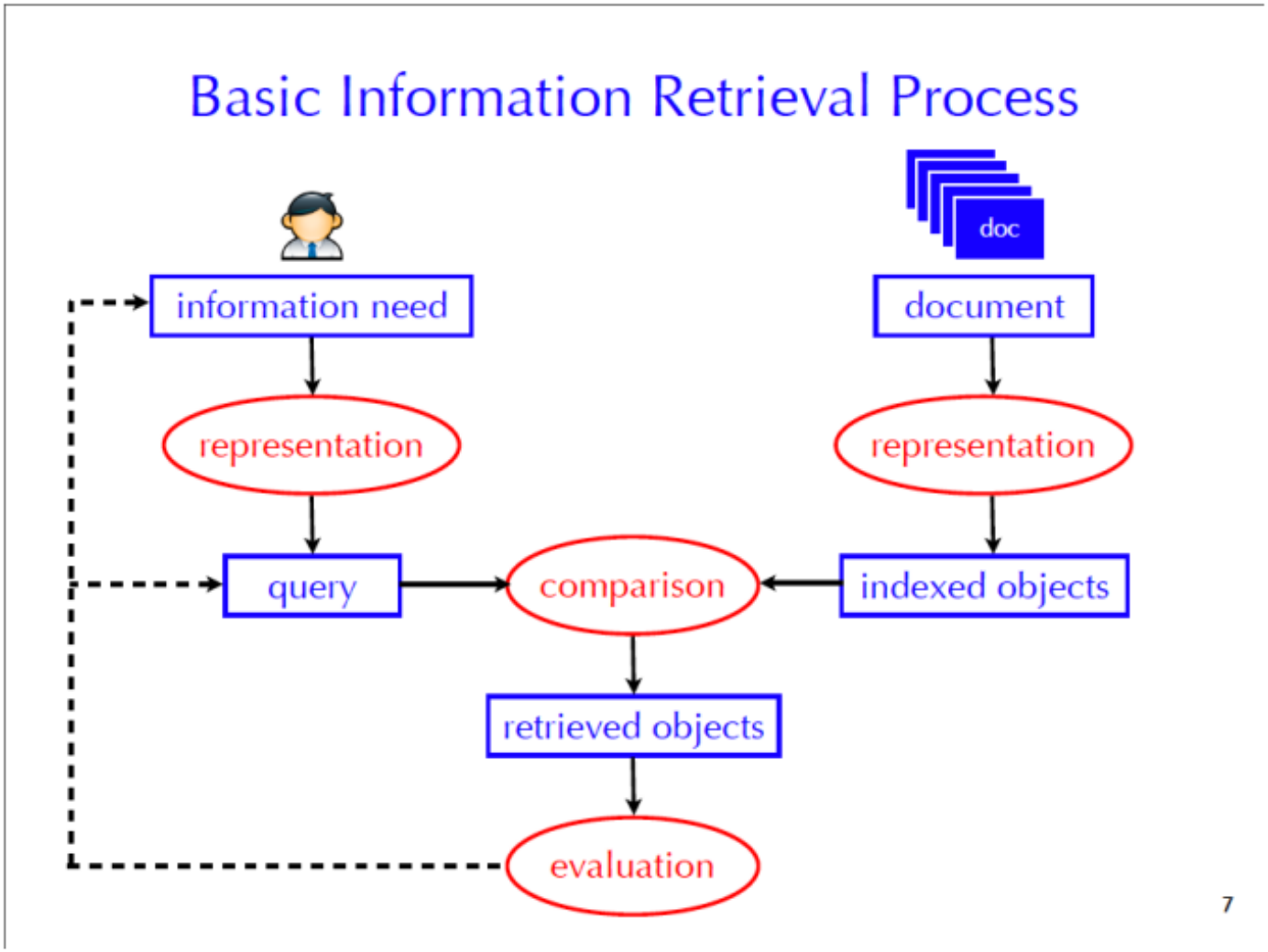


Diagram from Jaime Arguello, UNC-Chapel Hill

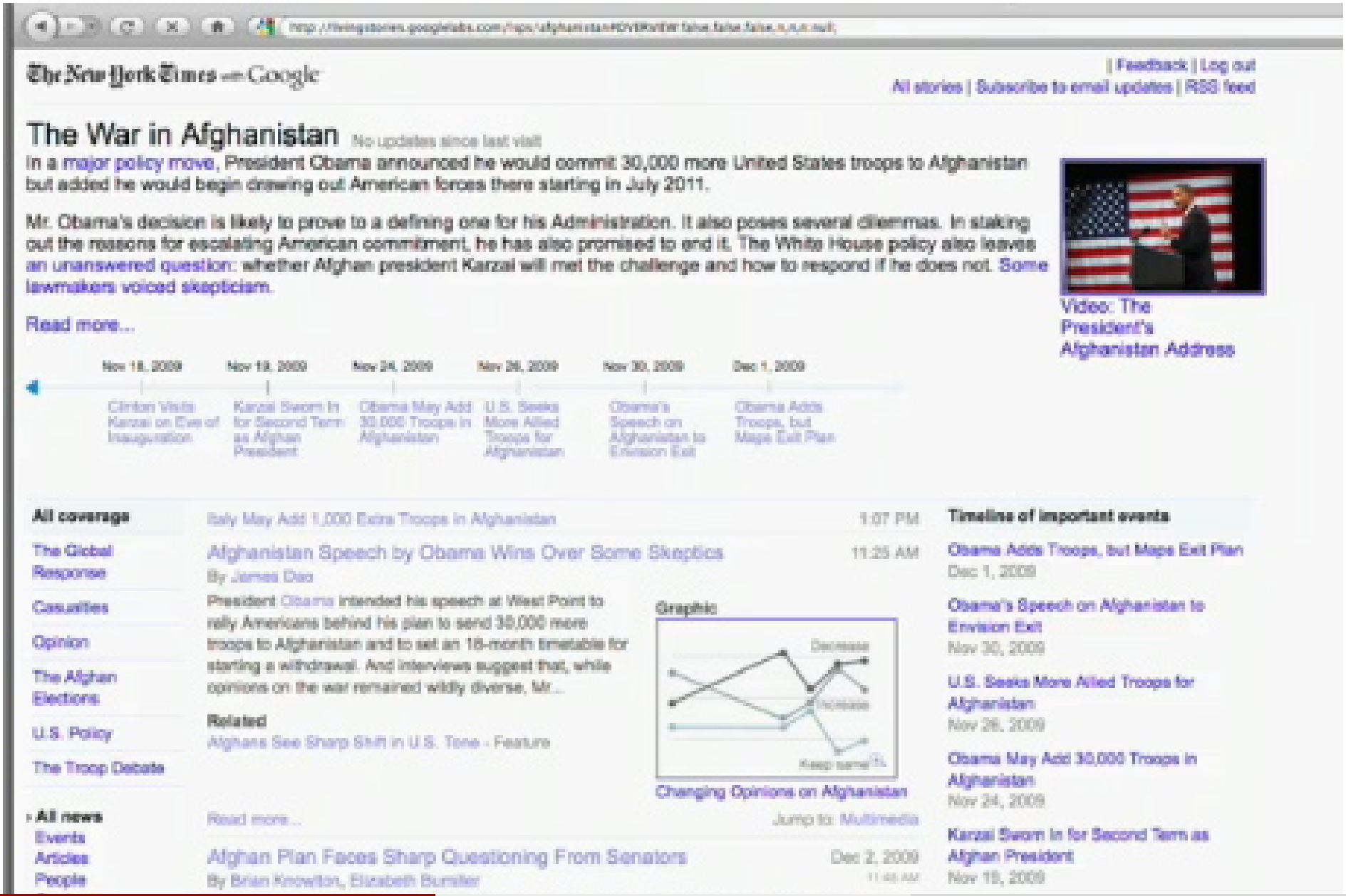
Text mining

Text mining (Hearst 2003)

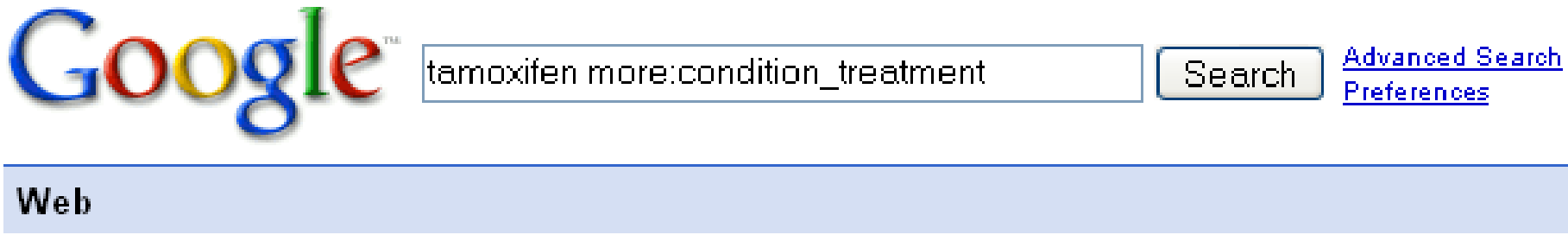
Text mining (TM) is the discovery by computer of new, previously unknown information, by automatically extracting information from different written resources. A key element is the linking together of the extracted information to form new facts or new hypotheses to be explored further by more conventional means of experimentation.

- Different from IR because the goal is to discover unknown information, which could not have yet been written down
- Different from data mining because TM extracts patterns from natural language texts rather than from structured databases of facts

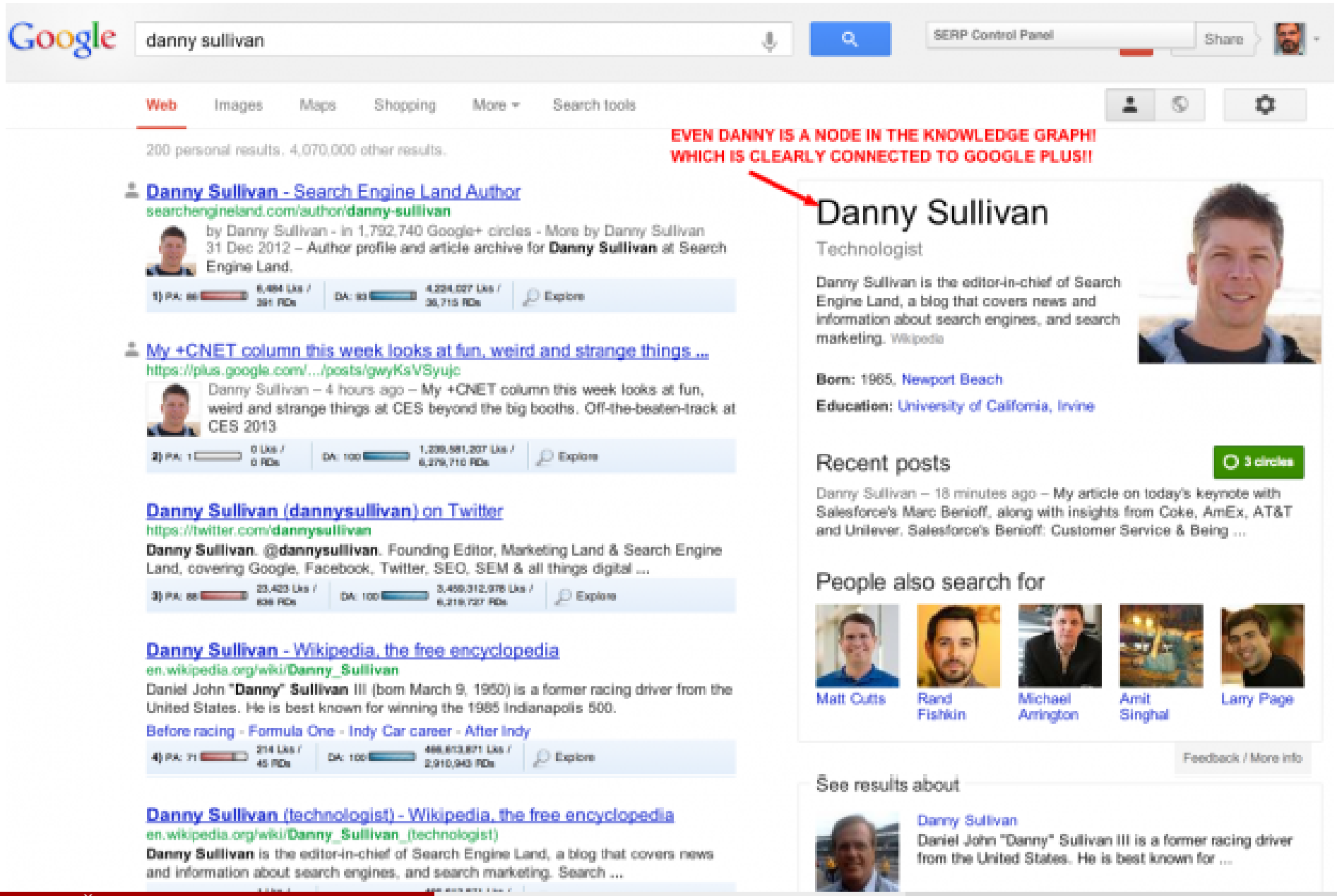
Text mining


The New York Times on Google
| Feedback | Log out
All stories | Subscribe to email updates | RSS feed
The War in Afghanistan
No updates since last visit
In a major policy move, President Obama announced he would commit 30,000 more United States troops to Afghanistan but added he would begin drawing out American forces there starting in July 2011.
Mr. Obama's decision is likely to prove to a defining one for his Administration. It also poses several dilemmas. In staking out the reasons for escalating American commitment, he has also promised to end it. The White House policy also leaves an unanswered question: whether Afghan president Karzai will met the challenge and how to respond if he does not. Some lawmakers voiced skepticism.
Read more...
Video: The President's Afghanistan Address
Nov 18, 2009
Nov 19, 2009
Nov 24, 2009
Nov 26, 2009
Nov 30, 2009
Dec 1, 2009
Clinton Visits Karzai on Eve of Inauguration
Karzai Sworn In for Second Term as Afghan President
Obama May Add 30,000 Troops in Afghanistan
U.S. Seeks More Allied Troops for Afghanistan
Obama's Speech on Afghanistan to Envision Exit
Obama Adds Troops, but Maps Exit Plan
All coverage
The Global Response
Casualties
Opinion
The Afghan Elections
U.S. Policy
The Troop Debate
All news
Events
Articles
People
Italy May Add 1,000 Extra Troops in Afghanistan
1:07 PM
Afghanistan Speech by Obama Wins Over Some Skeptics
11:25 AM
By James Dao
President Obama intended his speech at West Point to rally Americans behind his plan to send 30,000 more troops to Afghanistan and to set an 18-month timetable for starting a withdrawal. And interviews suggest that, while opinions on the war remained wildly diverse, Mr...
Related
Afghans See Sharp Shift in U.S. Tone - Feature
Graphic
Decrease
Increase
Keep same
Changing Opinions on Afghanistan
Read more...
Jump to: Multimedia
Afghan Plan Faces Sharp Questioning From Senators
Dec 2, 2009
By Brian Knowlton, Elizabeth Bumiller
Timeline of important events
Obama Adds Troops, but Maps Exit Plan
Dec 1, 2009
Obama's Speech on Afghanistan to Envision Exit
Nov 30, 2009
U.S. Seeks More Allied Troops for Afghanistan
Nov 26, 2009
Obama May Add 30,000 Troops in Afghanistan
Nov 24, 2009
Karzai Sworn In for Second Term as Afghan President
Nov 19, 2009

Information retrieval + Text mining



Information retrieval + Text mining



Unstructured data = text ?

Unstructured data (Manning and Schütze 2008)

Data that does not have clear, semantically overt, easy-for-a-computer structure. It is the opposite of structured data, the canonical example of which is a relational database.

- NB: In reality, almost no data are truly unstructured
- Text does have structure: linguistic structure (syntax, discourse, etc.) and formatting structure (headings, paragraphs, etc.)

Why is TAR a challenge?

- Text is written in natural language. Natural language is a tough nut to crack. It is complex, ambiguous, vague, and relies on commonsense knowledge.
- Full understanding of natural language is an AI-complete problem.
- On top of this: dealing with large amounts of data poses serious technical challenges.

Language complexity

- Phonology – relations between words and sounds made to realize them (not in purview of TAR)
- Morphology – shape and behavior of words in context
- Syntax – ordering and grouping of words
- Semantics – meaning of words (lexical semantics) and larger units (compositional semantics)
- Pragmatics – how language is used to accomplish goals
- Discourse – how text is structured into larger units

- World knowledge

Language vagueness

Rezultati pretrage za *mali stan*

1 2 3 4 5 6 7-12 SLJEDEĆA »

6091 oglas

Sortiraj Relevantnosti

Vau Vau Njuškalo oglasi

Srdoči, odličan mali stan u novogradnji

Stan u stambenoj zgradi, Prizemlje
Stambena površina: 24.44 m2

Objavljen: 03.03.2014.

46.500 € ~ **355.685 kn**

KVATERNIKOV TRG - UREĐEN mali stan

Stan u stambenoj zgradi, 5 kat
Stambena površina: 40.00 m2

Objavljen: 06.03.2014.

70.000 € ~ **535.441 kn**

MALI STAN U FAŽANI !!

Stan u stambenoj zgradi, ---
Stambena površina: 35.00 m2

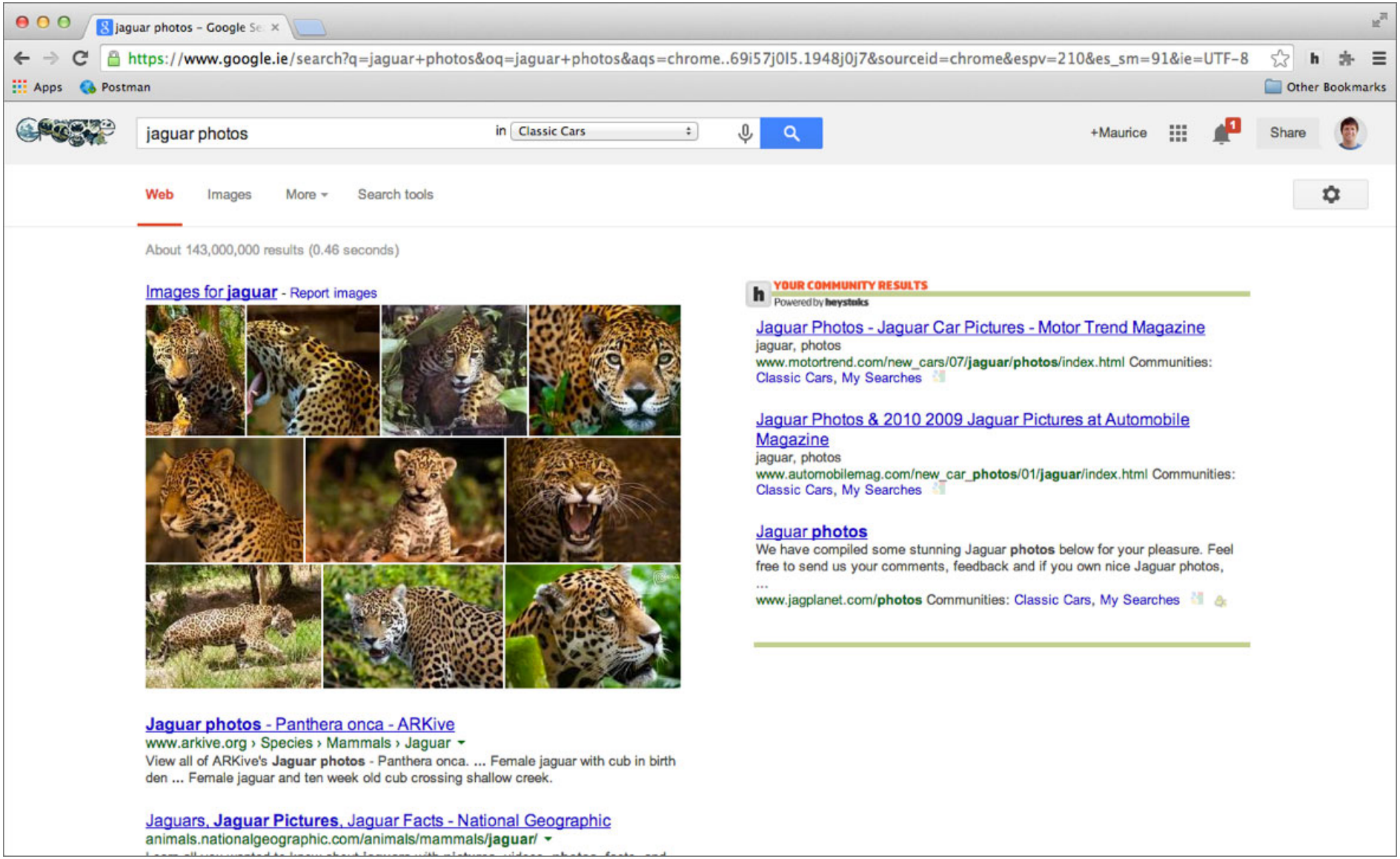
Objavljen: 05.03.2014.

48.500 € ~ **370.984 kn**

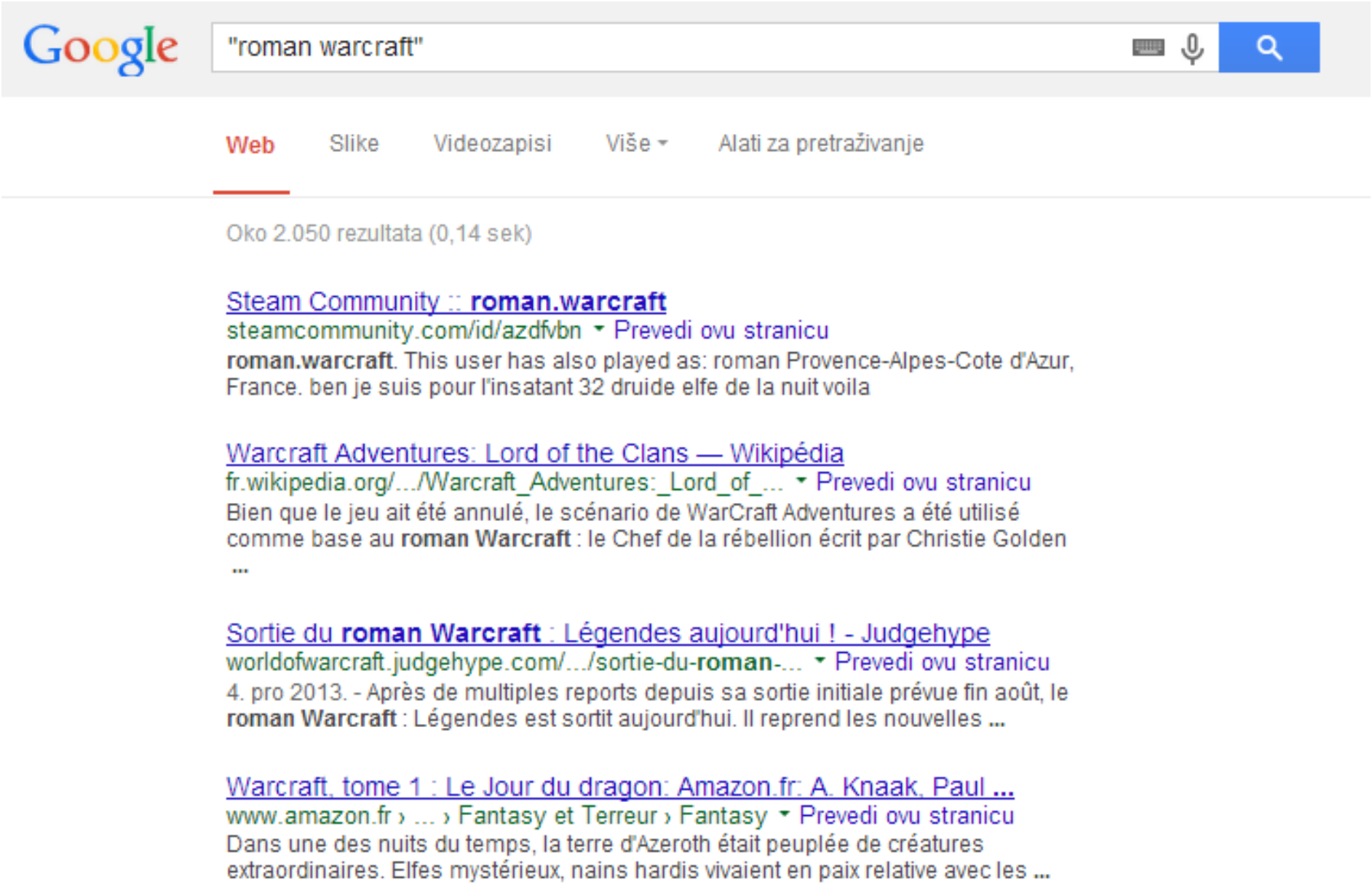
Language ambiguity

- Ambiguity increases the number of possible interpretations (combinatorial explosion of search space)
- Categorial ambiguity:
  - *Flying planes can be dangerous.*
  - *Time me on the last lap.*
- Word sense ambiguity:
  - *I saw her run to the bank.*
  - *The thief was charged by the police and had to pay a fine.*
- Structural ambiguity:
  - *I saw a boy on the hill with a telescope.*
- Referential ambiguity:
  - *Ann and Lisa gave John and Mark some apples because they liked them.*
  - *Lisa gave Ann a present and she said thanks.*

Word sense ambiguity



Word sense ambiguity



Natural language processing

Natural language processing *(Wikipedia)*

Natural language processing (NLP) is a field of computer science, artificial intelligence, and linguistics concerned with the interactions between computers and human (natural) languages. As such, NLP is related to the area of human–computer interaction.

- Morphological segmentation
- Part-of-speech tagging
- Parsing
- Word sense disambiguation
- Coreference resolution
- Named entity recognition
- Machine translation
- ...

Natural language processing

- Most text analysis subtasks addressed by NLP are useful for TAR
- Two main NLP paradigms:
  - Rule-based NLP (knowledge-based, symbolic, logic-based): state machines, formal rule systems, logic
  - Statistical NLP (non-symbolic): probability theory, statistics, machine learning
- The border is fuzzy
- Rule-based NLP is adequate for some tasks but providing exact and complete models is difficult ("All grammars leak"; Sapir 1921)
- Statistical NLP addresses this by detecting patterns occurring in texts
- Much of statistical NLP relies on (statistical) machine learning

Course aims and target audience

Aims

Provide a systematic overview of both traditional and advanced methods for text analysis and retrieval. The course is divided into two parts. The first part covers the fundamentals: NLP, IR, and ML. The second part focuses on concrete applications in information extraction and text analytics, with an emphasis on methods based on statistical natural language processing and machine learning.

Audience

Students interested in gaining an understanding of and practical experience with basic information retrieval and text analysis techniques. Knowledge of at least basic machine learning is required. No previous knowledge of natural language processing is required.

What this course is NOT about

- Computational linguistics – we'll cover only as much as we need for TAR, but won't go into much depth (e.g., morphology, syntax)
- Machine learning – we'll use models that are useful for TAR, but won't explain the inner workings
- String searching/matching algorithms (finding patterns within texts) – not really related to TAR
- Multimedia search (search for images, video, etc.) – out of scope, we focus on text only
- Cross-lingual and multi-lingual IR – interesting topics, but out of scope

Outline

1 Why this course?

2 Topics

3 Organization

Topics

1. Introduction
2. Basics of NLP: pipeline, parsing, copora & language models
3. Basics IR: models, evaluation, web search
4. Machine learning for NLP: tasks, sequence labeling, annotation
5. Semantics: lexical semantics, semantic parsing, distributional semantics
6. Neural NLP: RNNs, attention, contextualized embeddings
7. Recap & Prep

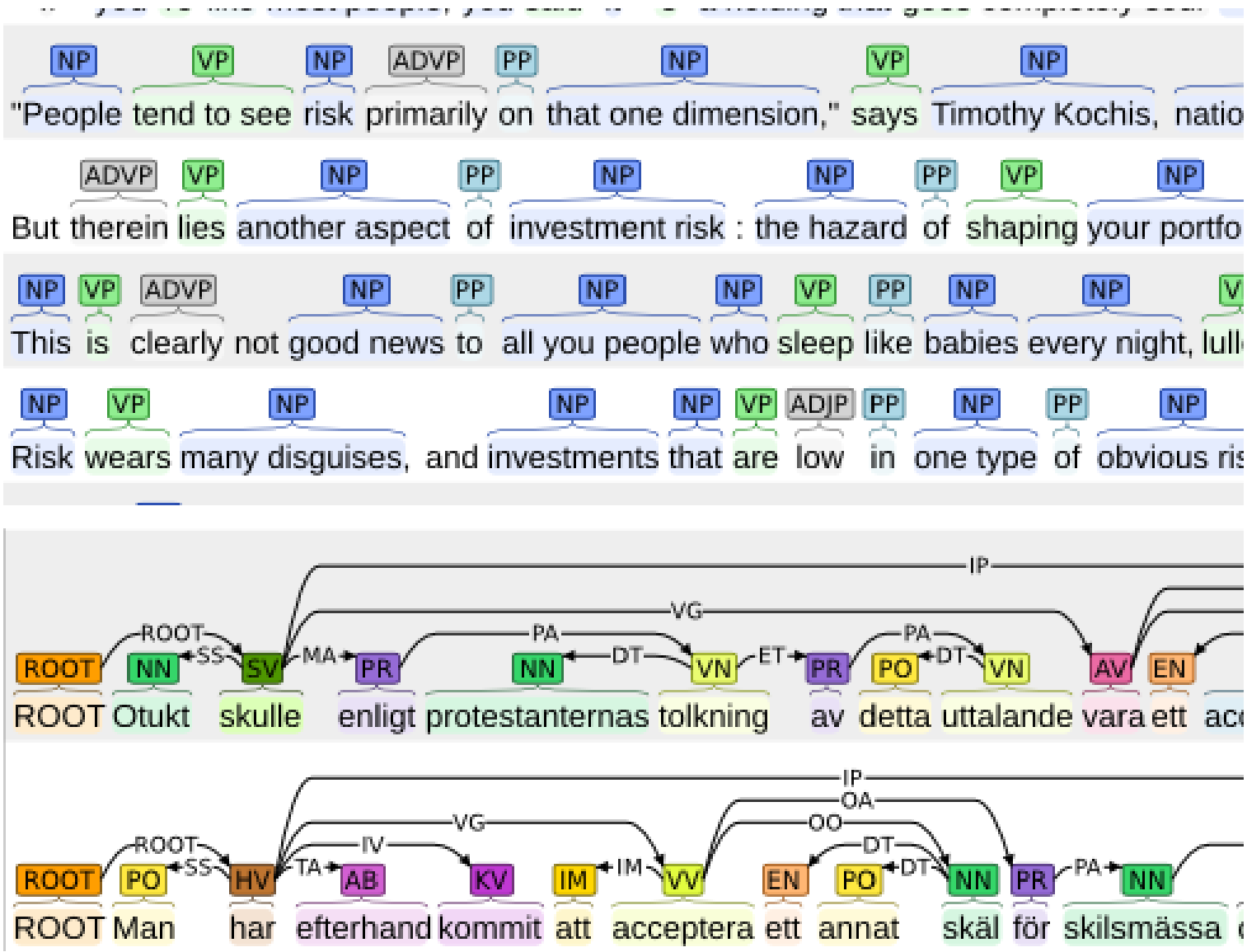
   (two weeks break)
8. Information extraction
9. Question answering
10. Text similarity & natural language inference
11. Text summarization & simplification
12. Sentiment analysis
13. Author profiling; Wrap-up

NLP and ML techniques

- Relevant NLP tasks:
  - Stemming and lemmatization
  - Language modeling and POS-tagging
  - Parsing
  - Word sense disambiguation
- Relevant ML techniques:
  - Classifiers: NB, logistic regression, MaxEnt, SVM
  - Sequence labeling: HMM, CRF
  - Clustering
  - Model selection and evaluation
- Available tools, libraries, and resources

Parsing

Chunking and parsing with Brat (`http://brat.nlplab.org`)



Basic IR models

- Text preprocessing
- Models
  - Boolean retrieval
  - Vector space model
  - Language model and smoothing
  - Probabilistic model
- Data structures for IR
- Available tools and platforms

Web search, advanced IR, and IR evaluation

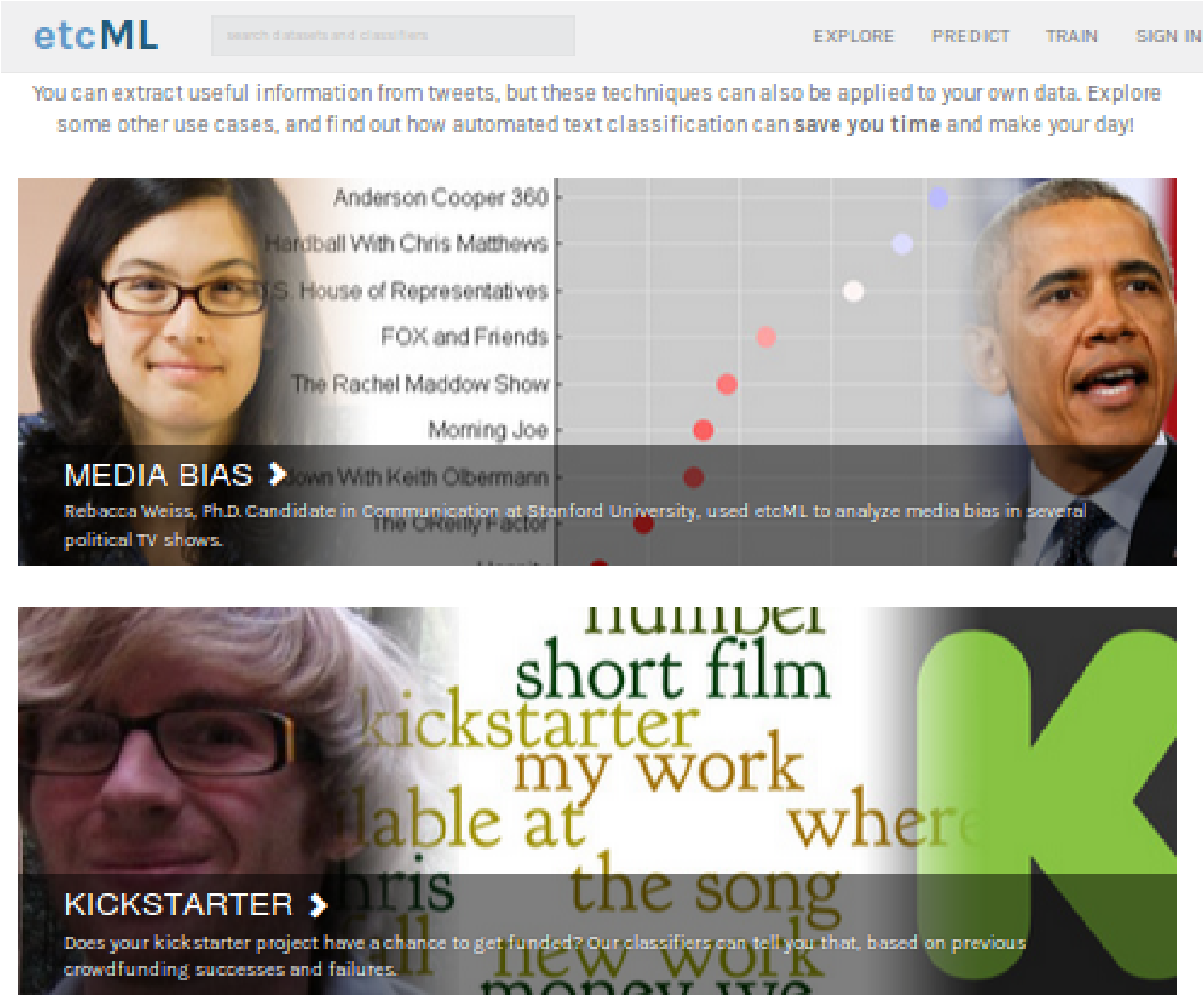
- Web search
  - HITS
  - PageRank
- Advanced IR
  - Query expansion
  - Relevance feedback
  - Semantic search
  - Topic detection and tracking
- Evaluation
  - Retrieval collections
  - Evaluation measures
  - TREC and CLEF evaluation campaigns

Text classification and clustering

- Text classification
  - Classification algorithms
  - Feature selection
  - Hierarchical document classification
  - Multi-label document classification
  - Evaluation
- Document clustering
  - Document similarity measures
  - Clustering algorithms
  - Evaluation
- Available tools and libraries

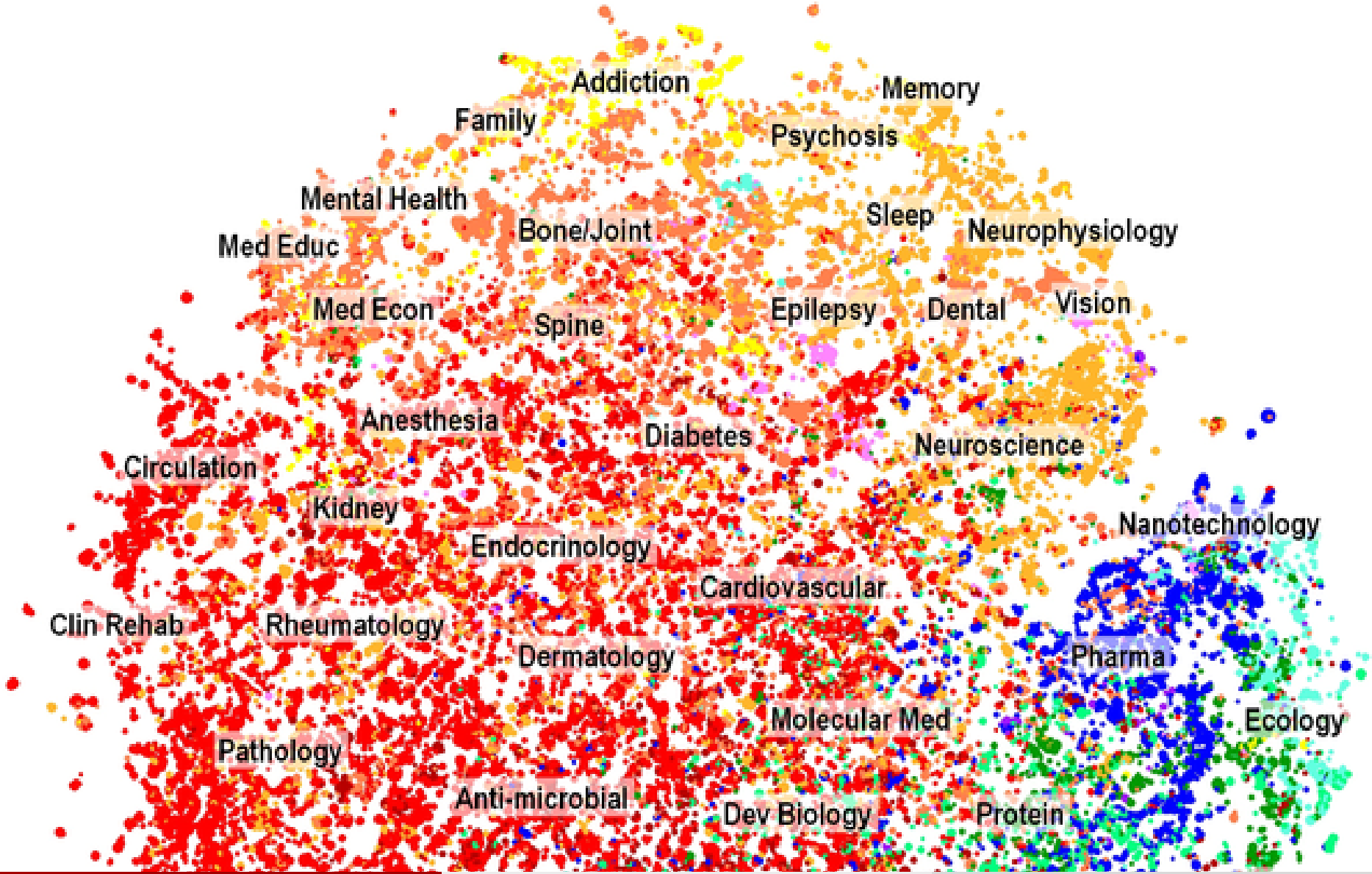
Text classification

Easy Text Classification (`http://www.etcml.com/`)



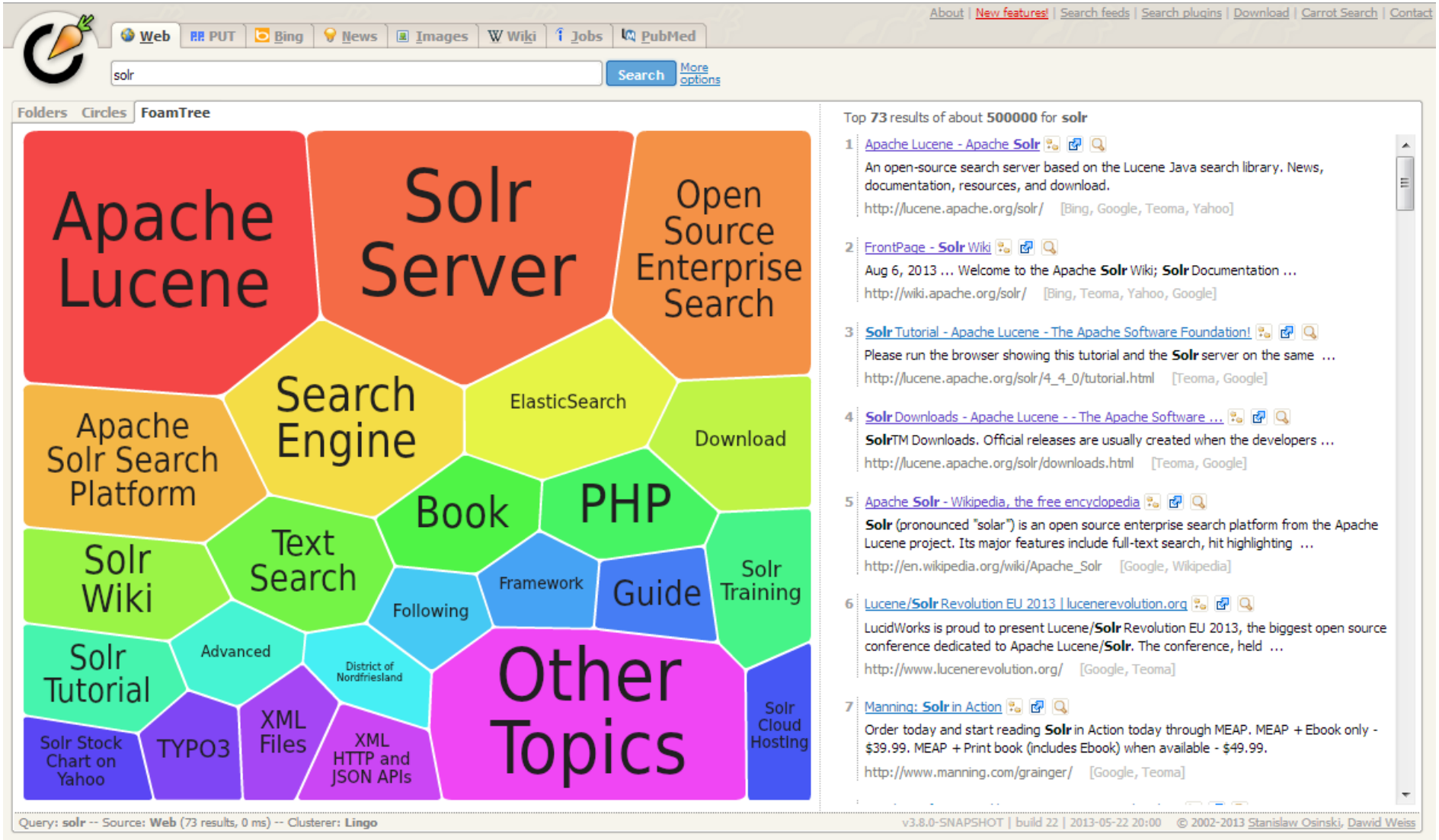
Document clustering

Clustering biomedical publications (http://tinyurl.com/kyludsb)



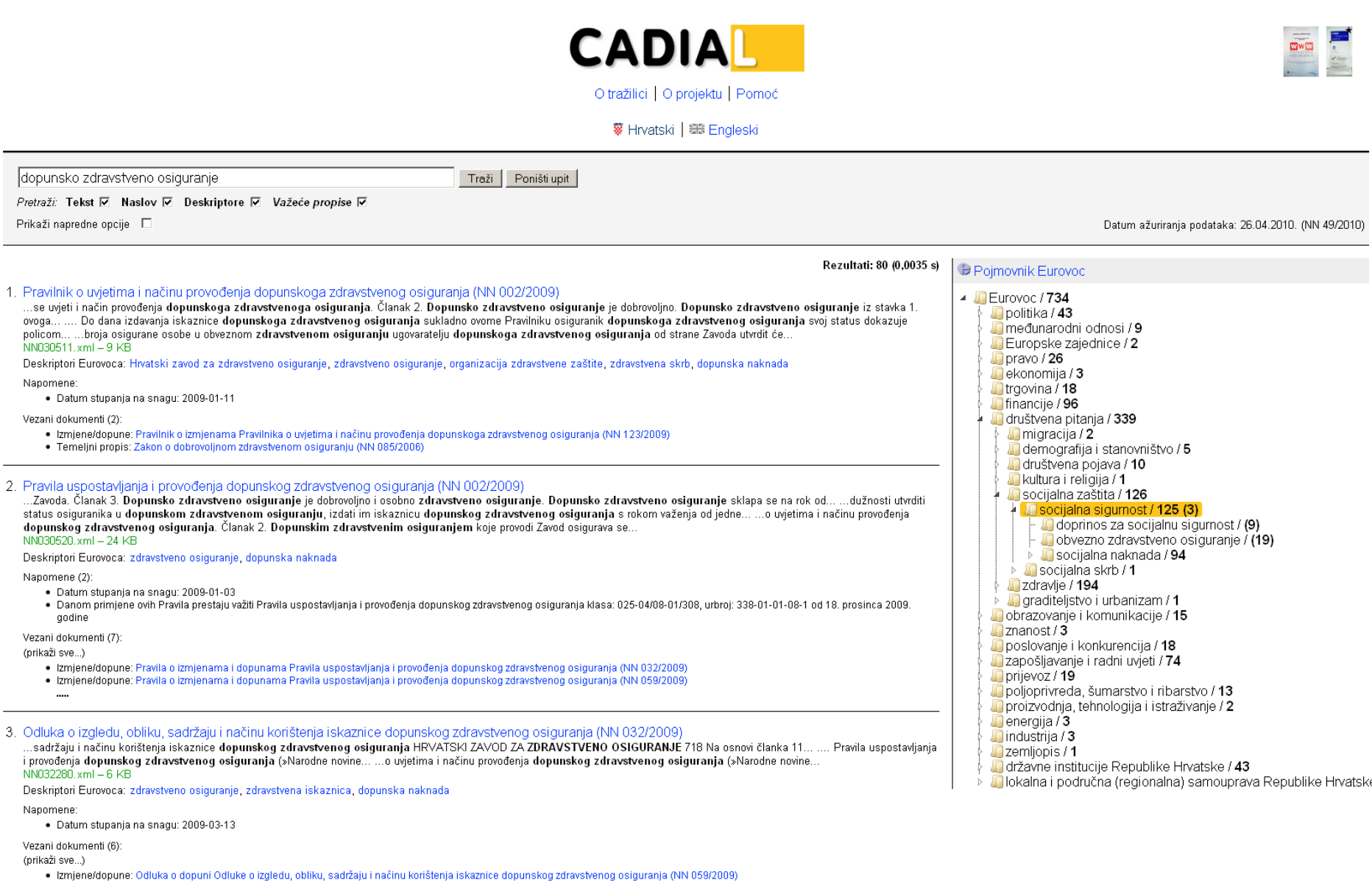
Faceted search

Carrot Search (`http://carrotsearch.com`)

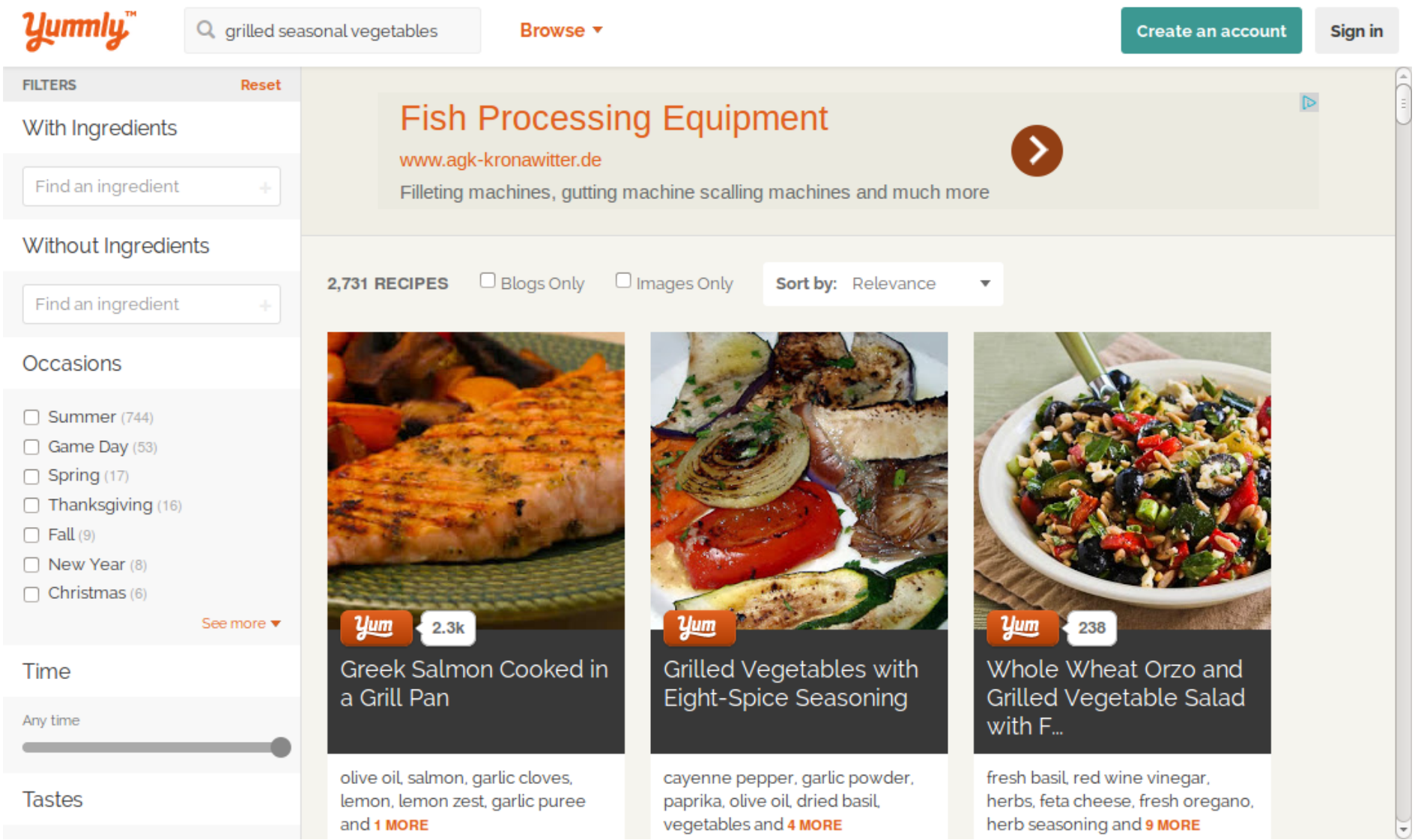


Faceted search

TakeLab CADIAL (`http://cadial.hidra.hr`)

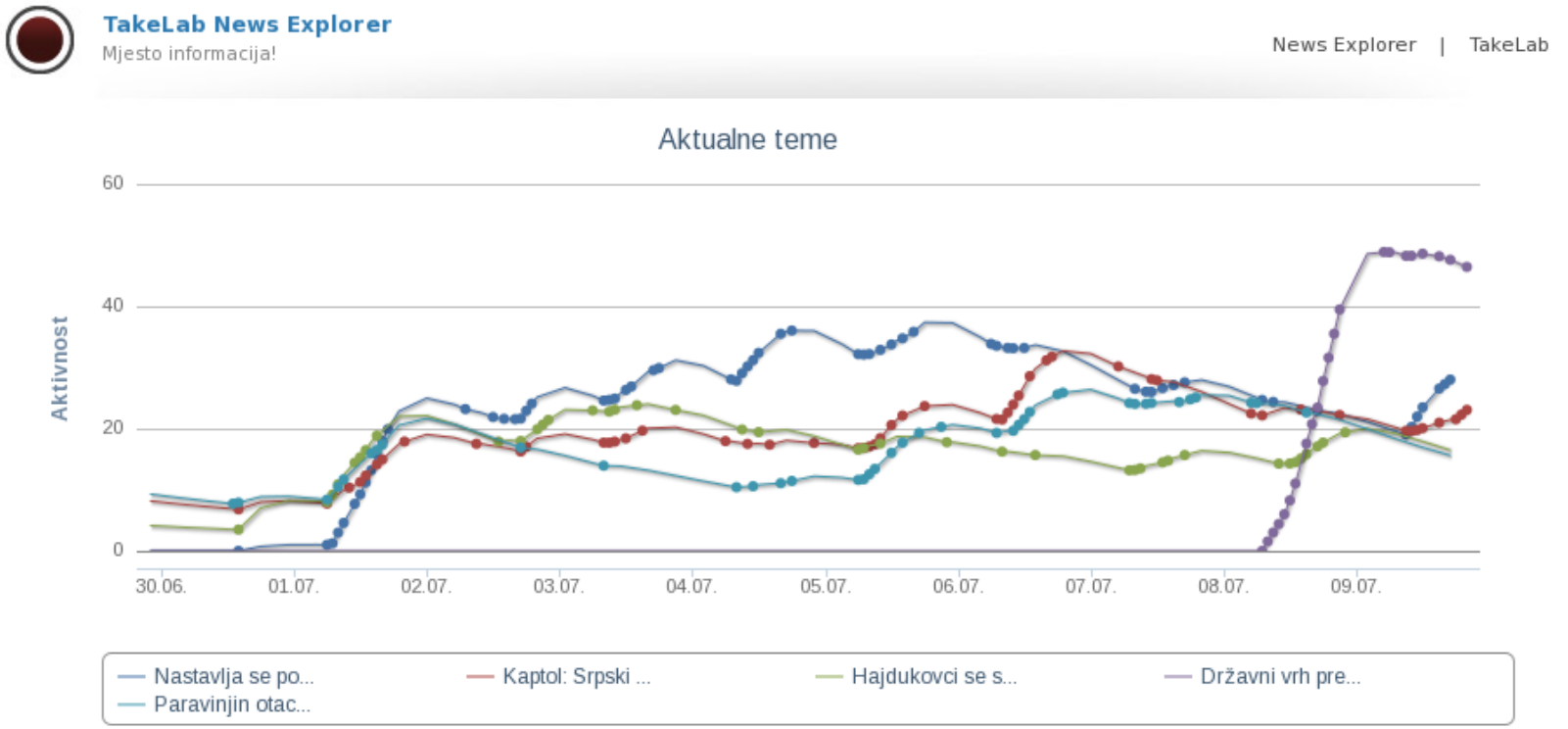


Semantic search



Topic detection and tracking

TakeLab News Explorer (`http://takelab.fer.hr/ne`)



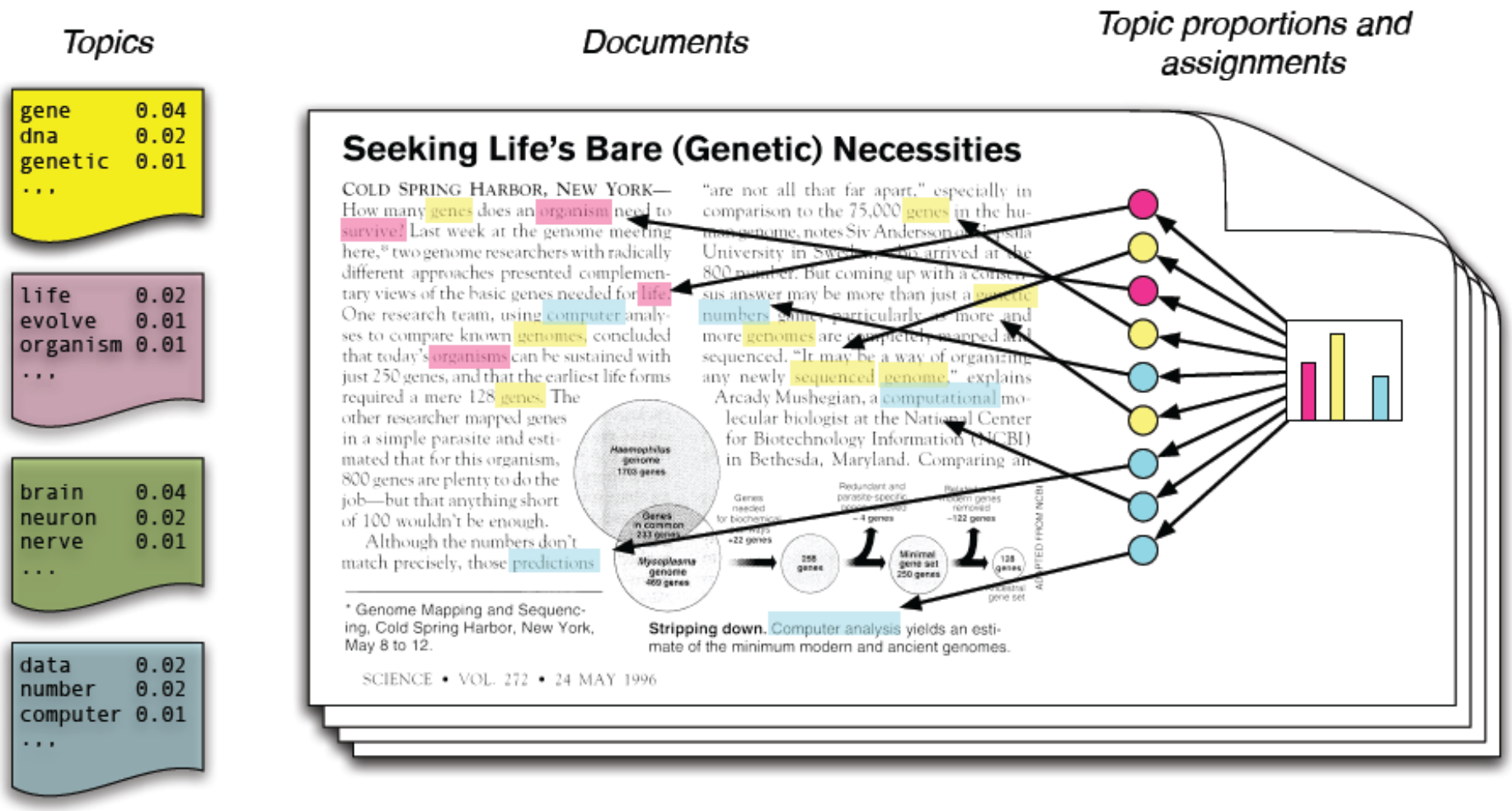
O tome se piše	Broj članaka
Nastavlja se potraga za Antonijom Bilić, ispitat će se i svjed ...	108
Kaptol: Srpski predsjednik uvrijedio nas je kada je aludirao d ...	79
Hajdukovci se spremaju u miru Postira, Nižetić tvrdi: Bodovi o ...	63
Državni vrh pretresa slučaj generala Kruljca	59
Paravinjin otac platio žrtvi da povuče optužbe?	58

O njima se piše	Broj članaka
Jadranka Kosor	286
Antonije Bilić	247
Dragana Paravinje	211
Ivo Sanader	122
Željko Kerum	95
Ivo Josipović	90

Latent document models

- Latent semantic indexing
- Probabilistic latent semantic indexing
- Topic models (Latent Dirichlet allocation)

Latent document models

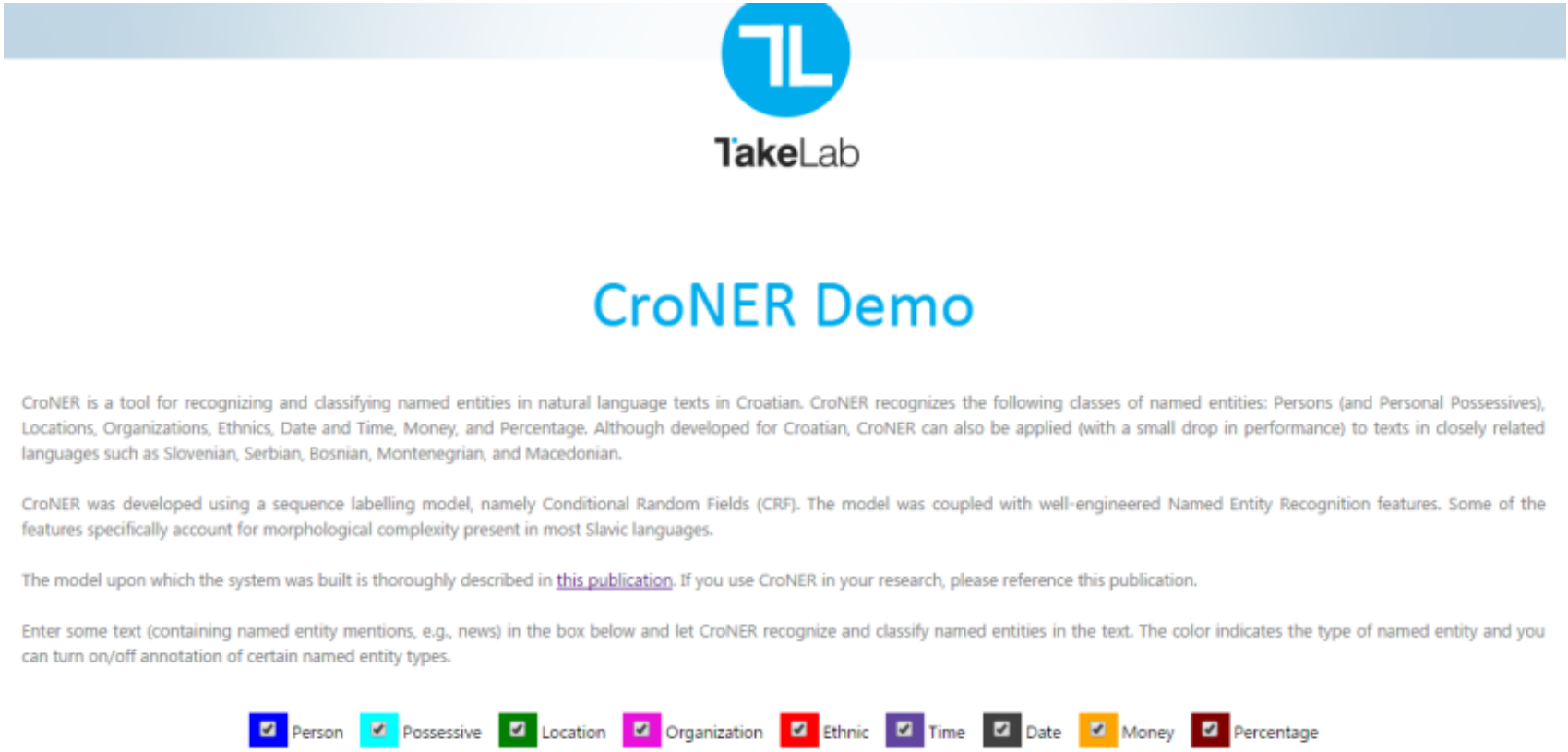


Information extraction

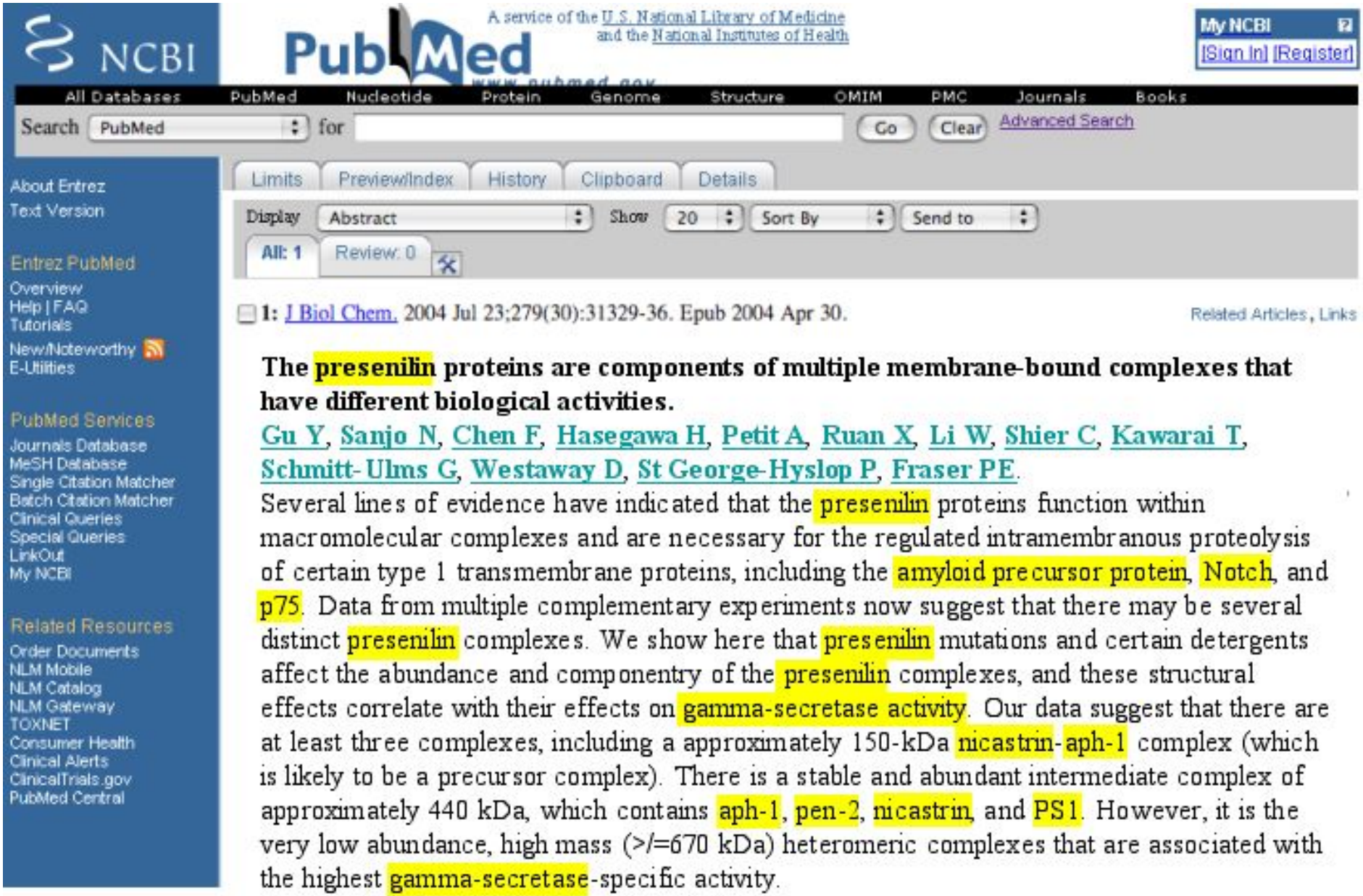
- Keyphrase extraction
- Named entities extraction
- Relation extraction
- Event extraction
- Temporal text processing

Named entity extraction

TakeLab CroNER (`http://takelab.fer.hr/CroNER`)



Named entity extraction



Named entity extraction

TakeLab MINERAL (`http://takelab.fer.hr/mineral`)

MINERAL: Clinical Information Extraction

Clinical narratives are abundant with mentions of diseases, disorders, and other clinical conditions. Many surface forms represent the same condition. In biomedicine, there are rich lexical and ontological resources that can be leveraged when building applications. The Unified Medical Language System (UMLS) represents over 130 lexicons/thesauri with terms from a variety of languages.

This is a Web Demo of MINERAL, TakeLab's system for identifying and normalizing disorders in clinical narratives. MINERAL was built for participation in the SemEval 2015 Task 14 Clinical information extraction. Console version of MINERAL, which can be used for bulk processing of large number of clinical documents is freely available here (see the README file for instructions on prerequisites and how to run the tool).

If you use MINERAL in your research, please reference the this publication (bibtex).

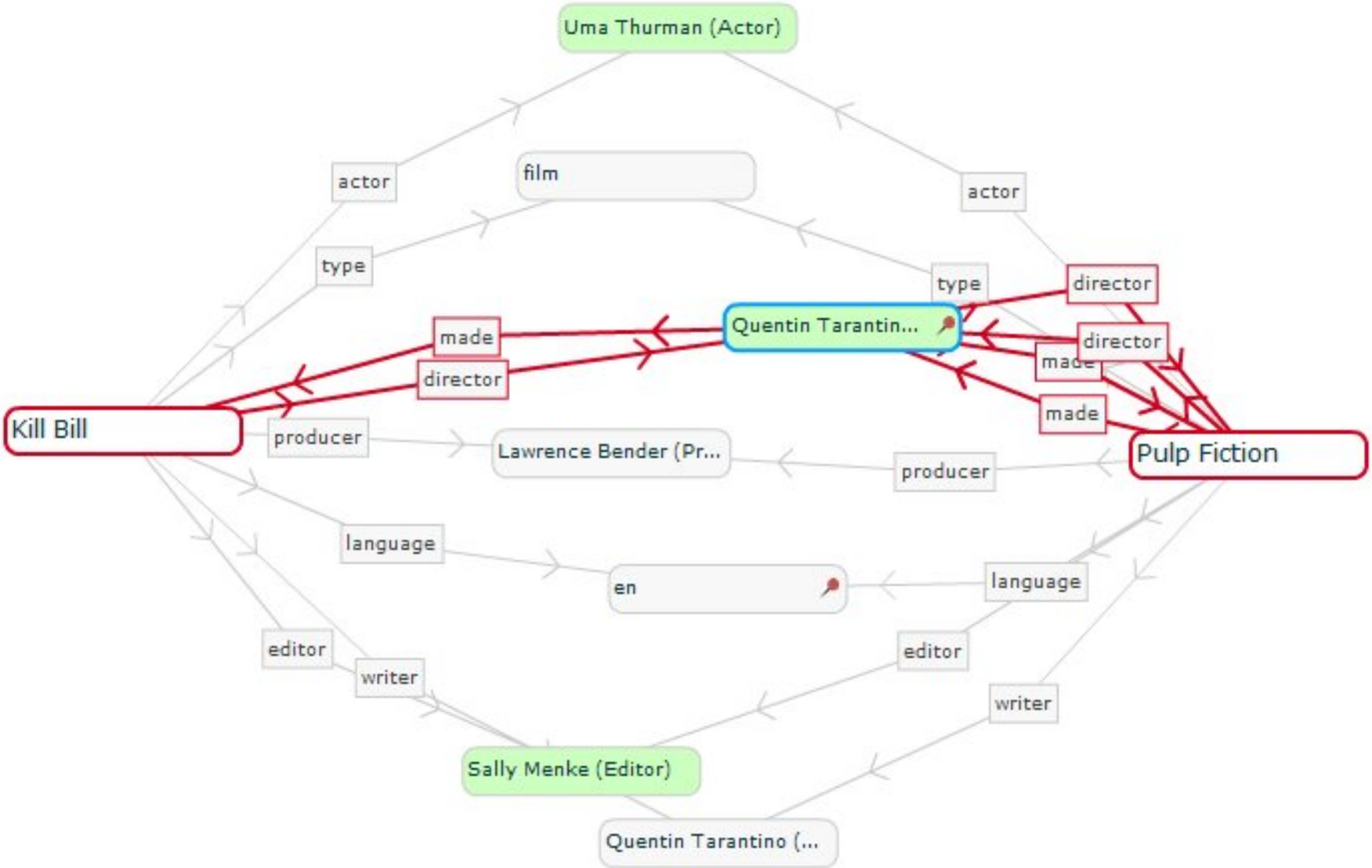
Enter some clinical/medical text in the box below and let MINERAL annotate disorder mentions (annotated as hyperlinks). By hovering over recognized diseases you can see the Concept Unique Identifier (CUI) and the standard name recognized disease has in the SNOMED-CT database.

The patient is an (XX)-year-old female with history of previous stroke; hypertension; COP Extracted mention: COPD
UMLS CUI: C0024117
SNOMED-CT name: chronic obstructive pulmonary disease finding
all and possible syncope. While walking, she accidentally fell to her knees and did hit her head ot observed, but the patient does not profess any loss of consciousness, recalling the entire event. The patient does have a history of previous falls, one of which resulted in a hip fracture. She has had physical therapy and recovered completely from that. Initial examination showed bruising around the left eye, normal lung examination, normal heart examination, normal neurologic function with a baseline decreased mobility of her left arm. The patient was admitted for evaluation of her fall and to rule out syncope and possible stroke with her positive histories.

CT CHEST WITH IV CONTRAST: Right-sided chest tube remains in place. There are bilateral pleural effusions, with persisting consolidation in the right lower lobe. The mid and upper lung zones are clear. The airway is patent to the segmental level. The pulmonary arteries enhance normally without filling defect. There are diffuse enlarged lymph nodes throughout the mediastinum, suspicious for metastatic nodes. The thoracic aorta is normal in caliber and enhances normally. Multiple small subcentimeter axillary nodes.

Relation extraction

RelFinder (http://www.visualdataweb.org/relfinder.php)

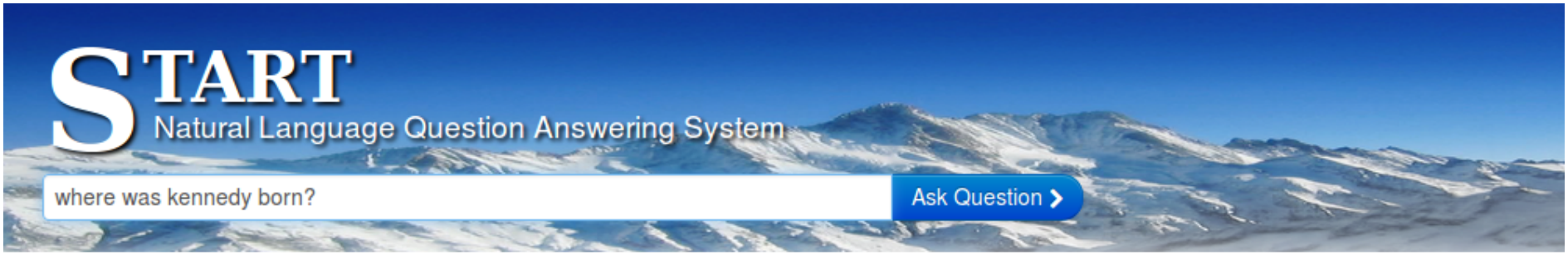


Question answering

- Closed-domain QA
- Open-domain QA
- Question classification
- QA evaluation

Question answering

START QA (`http://start.csail.mit.edu`)



===> where was kennedy born?

John Fitzgerald Kennedy was born May 29, 1917, in Brookline, Massachusetts
.

Source: Internet Public Library

Which person in The Internet Movie Database do you mean:

- Robert F. Kennedy
- Charles Kennedy (I)
- William Kennedy (II)

I know about 166 more persons called "Kennedy". You don't really want me to list them all, do you?

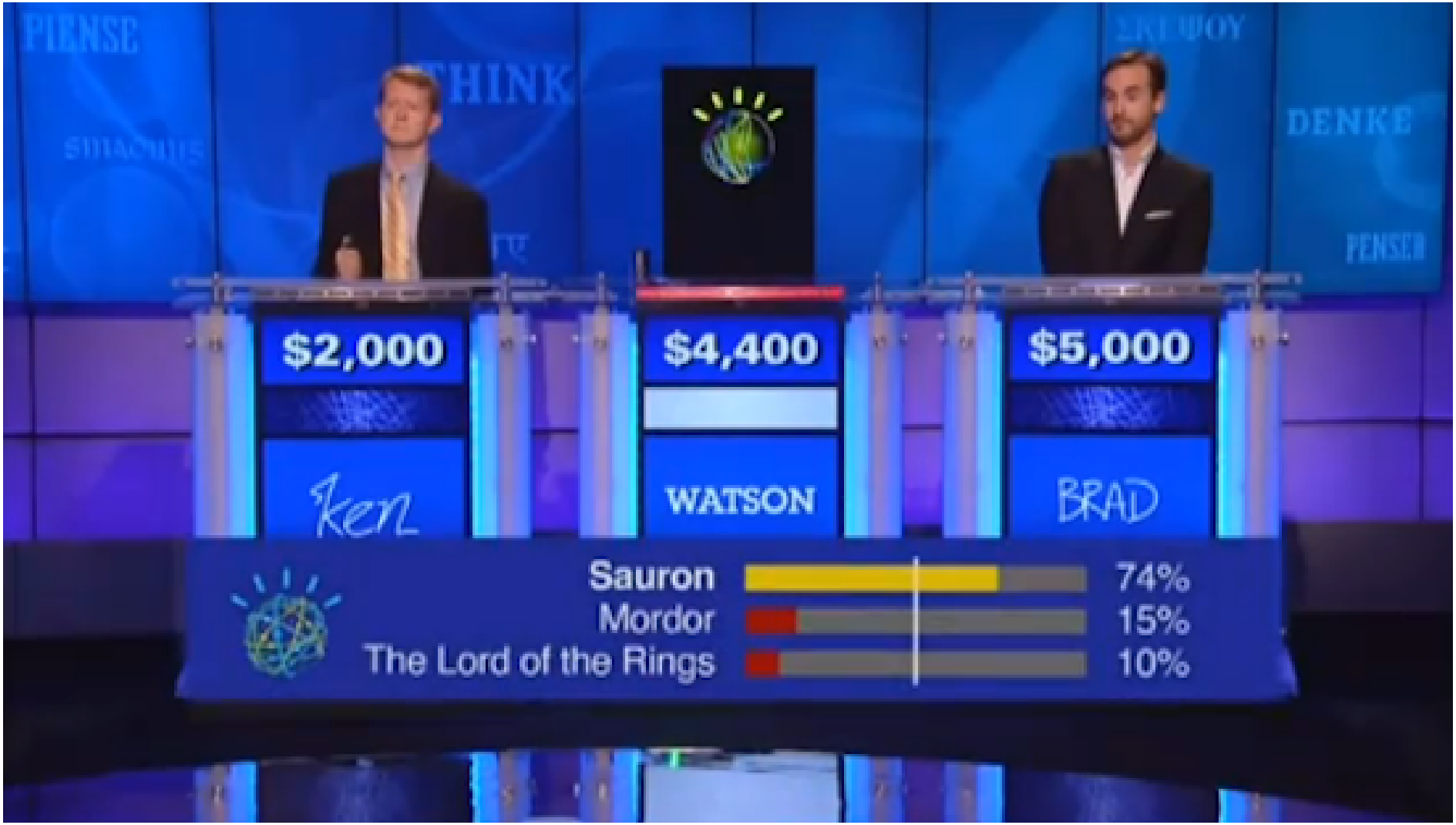
List them all

Which person in Wikipedia do you mean:

- 7166 Kennedy

Question answering

DeepQA project and IBM Watson



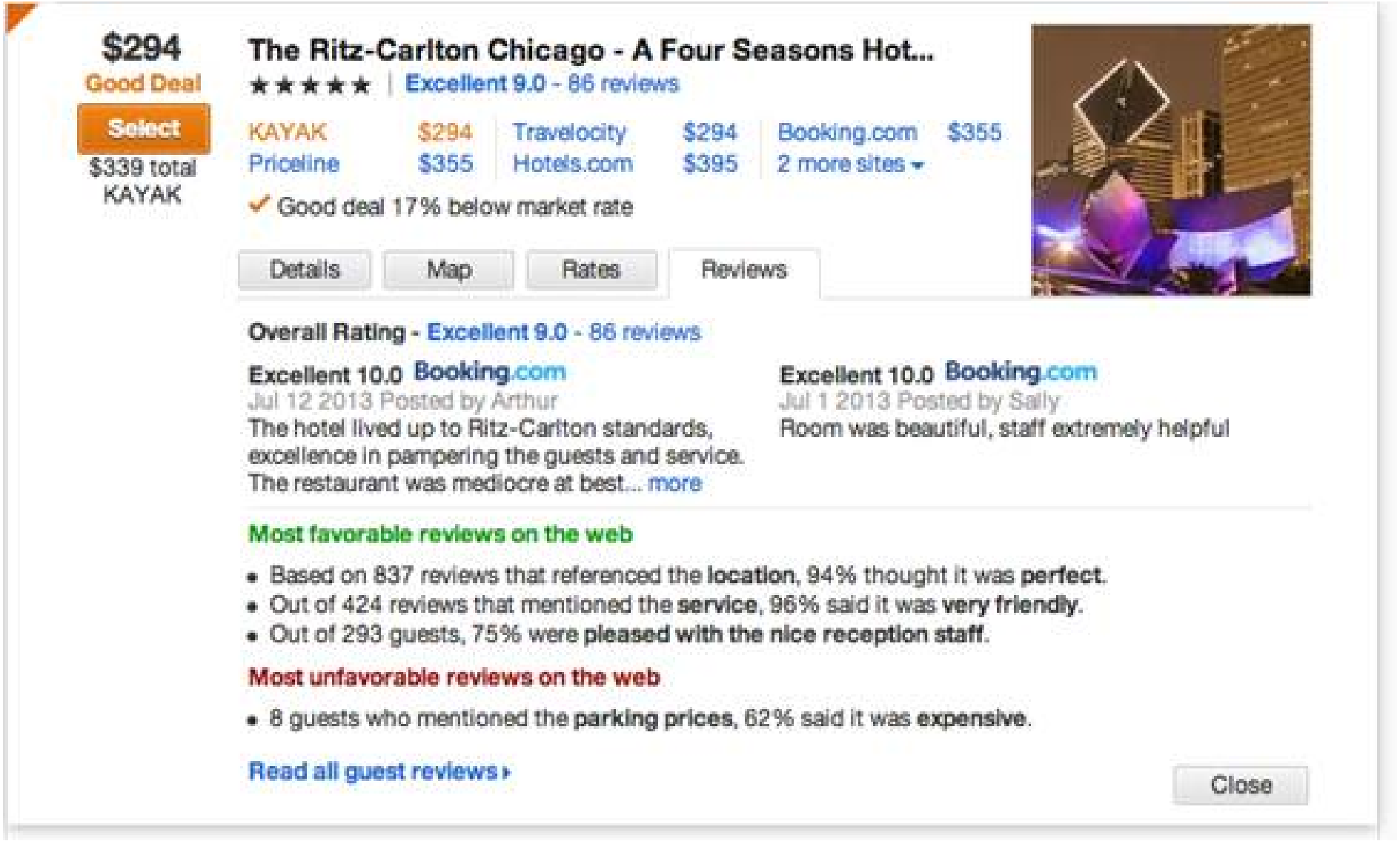
Text similarity, summarization, and simplification

- Semantic text similarity & paraphrase detection
- Textual entailment
- (Multi-)document summarization
- Text simplification

Sentiment analysis

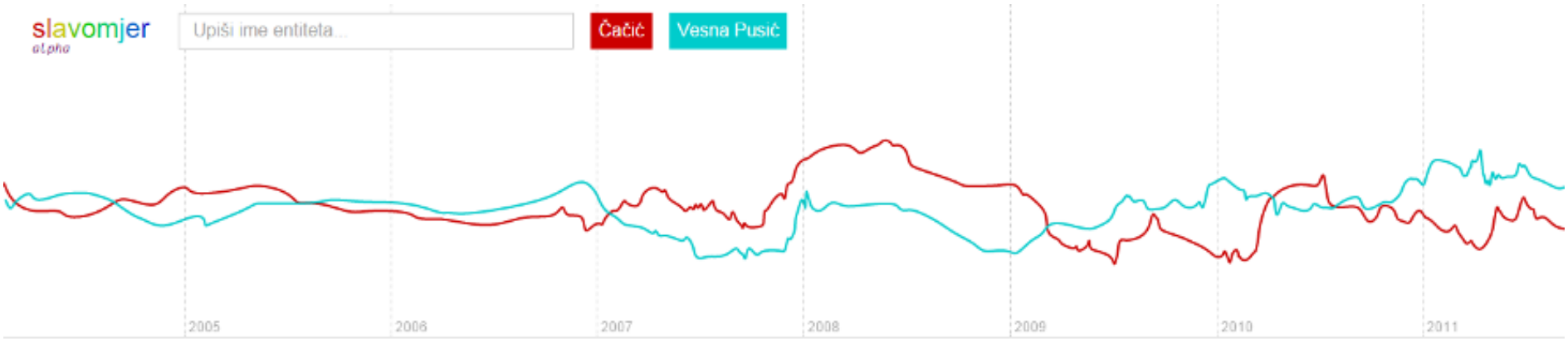
- Sentiment polarity detection
- Sentiment lexica
- Aspect-based sentiment analysis
- Sentiment summarization
- Available tools and resources

Sentiment analysis



Sentiment analysis

TakeLab Slavomjer (http://takelab.fer.hr/slavomjer)



Vesna Pusić +1 27.8.2014.
Ajde ne pričajte gluposti... · Ćirilici je mesto u svim Hrvatskim gradovima gde žive ili su nekad živeli Srbi,svidelo se to vama ili ne...

Vesna Pusić -1 27.8.2014.
Mene samo zanima jesu li kukurikavci stvarno takvi Jugoslaveni, jesu li toliko nesposobni ili je posrijedi nešto treće što još ne znamo ?

Vesna Pusić -1 26.8.2014.
Za sada jedinu razliku između Vesne Pusić i Mirele Holly vidim u javnom nastupu. · Dok je Vesna Pusić neodgojena, arogantna i bahata te to ni ne skriva, · Mirela više nastupa kao Josipović, · odmjereno, ukočeno, sa puno floskula a bez prave poruke i sadržaja.

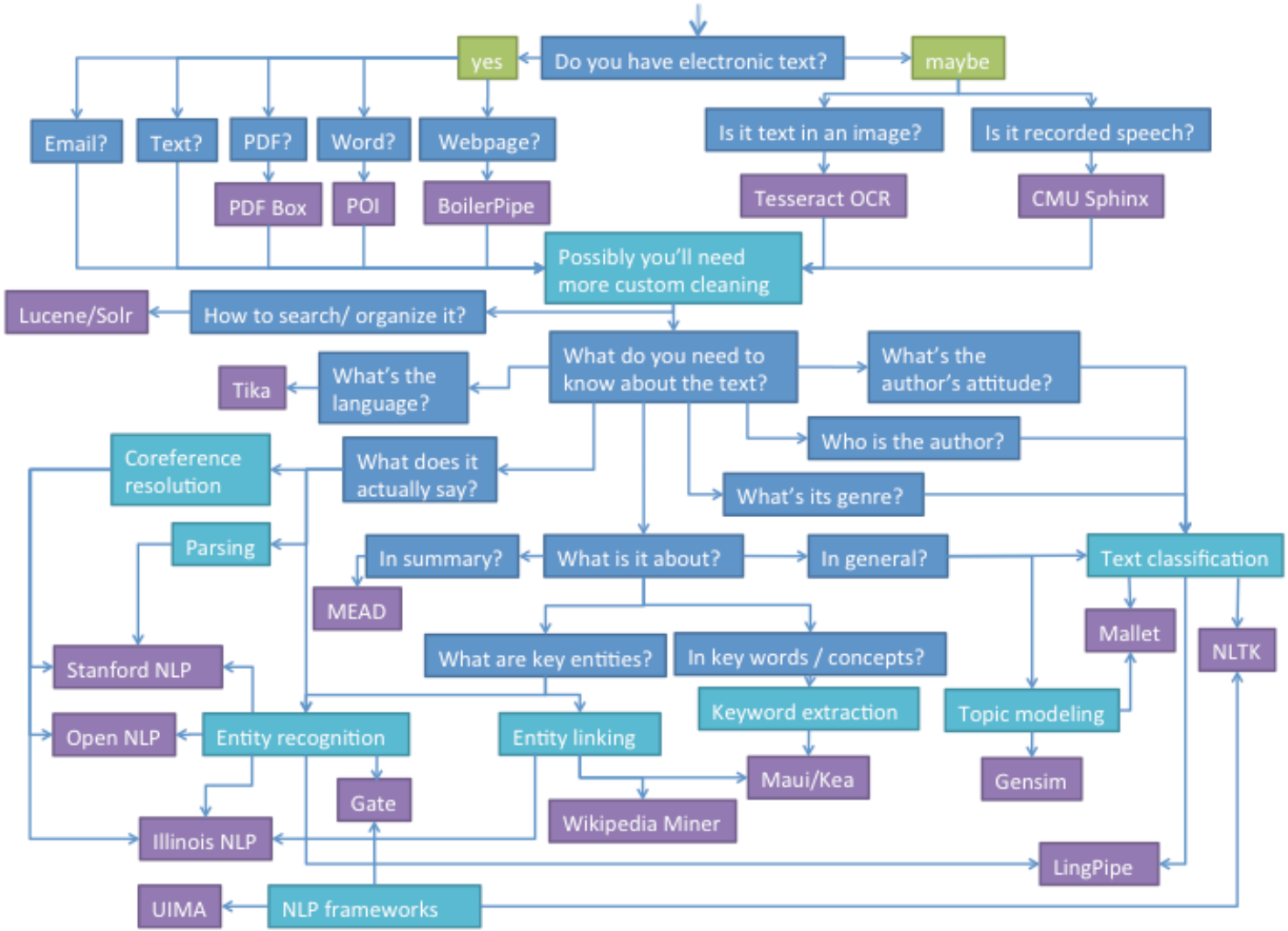
Vesna Pusić +1 19.8.2014.
To je izmislila tvoja draga Vesna Pusić, a ti samo sanjaj da će ti se ikad vratiti tvoja omiljena socrealistička vremena ali dočekati toga nećeš (a znam više proletera poput tebe)

Vesna Pusić +1 -2 17.8.2014.
Nažalost, Hrvatska je doslovno prodana, a EU je grob hrvatski, nema govora da unutar ovog groba bilo što krene na bolje, jednostavno ne može, na jesen opet stižu nove štednje i rezovi, cilj je da hrvatski radnik, po cijeni bude jednak onom u Bangladešu i Indiji. · Ako unutar EU-a ikad više budu veće plaće od ove Mrsićeve tri prajca, ja ću govno pojest pred svima! · Rezovi nisu sredstvo protiv recesije, rezovi su cilj koji je EU postavila našim političarima!

Vesna Pusić +1 -3 17.8.2014.
citala sam nocas nekaj malo s bolkovicevog i kuljisevog sajta (koji s am otkrila t ek jucer ·) pa mi je zapeo b io za pr imozak je dan tekst di je kuljis pisal o krizi i titu.. kao, da je veclm dljelom on bio taj koji je osmislio bio scenski nastup, ozracje i smjernice nas nesvrstanih jugica · pa sam bas pomislila bila kako danas nemamo apsolutno nikoga da poradi na nasem, camorac, stajlingu (·), kako smo bas lutke na koncu bez ikakvog okusa i mirisa · to je bas tuzno · istina, a '74. nas je posebno razguzil, necu mu to nikad oprostit

Čačić +3 6.8.2014.
Čačić je u intervjuu Indexu konačno potvrdio da desničari u Hrvatskoj moraju glasati za liberale ako žele da se provode ideje konzervativaca. Margaret Thatcher sebe je u mladosti smatrala liberalom (klasičnim naravno) i učlanila se u

Tools jungle



http://entopix.com/so-you-need-to-understand-language-data-open-source-nlp-software-can-help.html

Outline

1 Why this course?

2 Topics

3 Organization

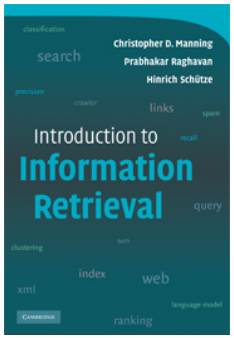
Learning outcomes

On successful completion of the course, you will be able to:

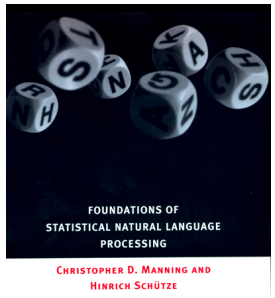
1. Summarize the application areas, trends, and challenges in text analysis and retrieval
2. Describe the fundamental techniques of text analysis and retrieval
3. Use linguistic preprocessing tools
4. Design and implement a text analysis/retrieval system
5. Apply machine learning algorithms to text analysis tasks
6. Evaluate a text analysis/retrieval system
7. Organize and formulate a system description paper
8. Describe, review, analyze, and criticize the main text analysis methods present in scientific papers

Textbooks

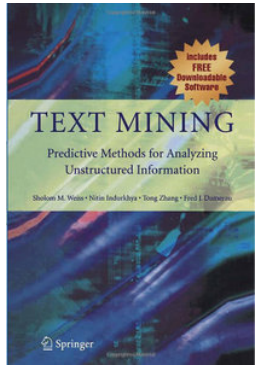
- C. D. Manning, P. Raghavan, H. Schütze. *Introduction to Information Retreval*, CUP, 2008



- C. D. Manning, H. Schütze: *Foundations of Statistical Natural Language Processing*, MIT Press, 1999

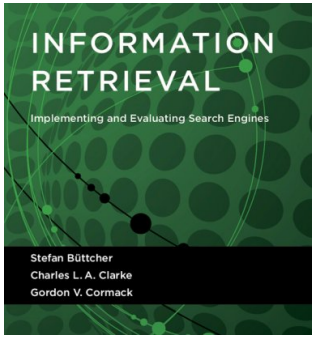


- S. M. Weiss, N. Indurkhya, T. Zhang, F. Damera. *Text Mining: Predictive Methods for Analyzing Unstructured Information*, Springer, 2010

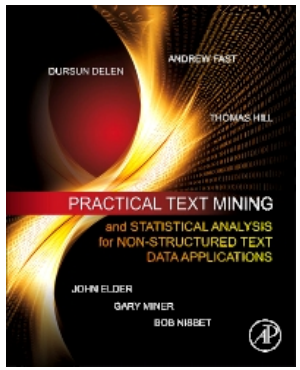


Textbooks

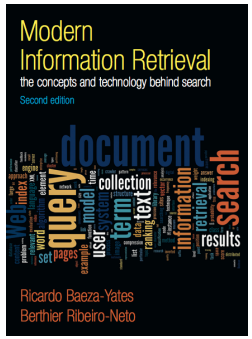
- S. Buettcher, C. L. A. Clarke, G. V. Cormack. *Information Retrieval: Implementing and Evaluating Search Engines*, MIT Press, 2010



- G. Miner, J. Elder IV, T. Hill, R. Nisbet, D. Delen, A. Fast. *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*, Academic Press, 2012

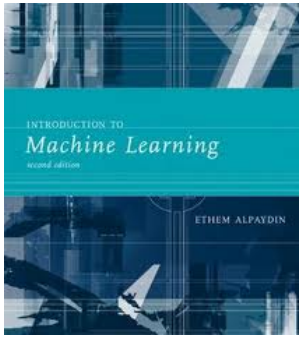


- R. Baeza-Yates, B. Ribeiro-Neto. *Modern Information Retrieval: The Concepts and Technology Behind Search*, ACM Press, 2011

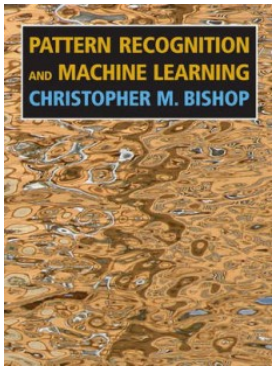


Textbooks – Machine learning

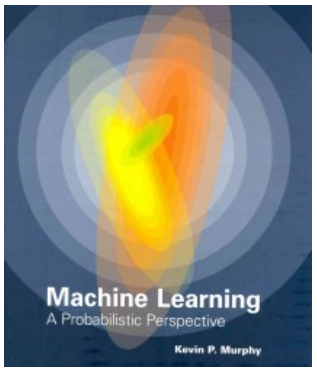
- Ethem Alpaydin: *Introduction to Machine Learning*, MIT Press, 2009



- Christopher Bishop: *Pattern Recognition and Machine Learning*, Springer, 2007



- Kevin P. Murphy: *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012



Exams

- Continuous assessment: no exams
  - Replaced by pre-reading and in-class quizzes (18% of grade) and paper reading (18% of grade)

Theory: pre-reading and in-class quizzes

- Replace the traditional midterm exam
- In the first part of the course (theory)
- Before each class, you'll have to read some materials (a book chapter or similar)
- In the class, you'll solve a quiz covering the pre-reading material (2 points) and last-week's material (4 points)
- Total of 30 points for 6 classes, scaled down to 18% of the grade
- A passing threshold of 15/30 points
- Designed to encourage continuous work
- Should be fairly simple for those who do follow this principle

Paper reading: in-class quizzes and discussions

- Replaces the traditional final exam
- You will be required to read a paper before the class and then we'll jointly discuss it in the class
- You'll be awarded 1 point for submitting your paper summary and your questions about the paper via an online form
- You'll get up to 5 points for the in-class quiz covering the paper (you should have the paper printout with you)
- Finally, you'll get 1 point for participating in the discussion
- Total of 35 points for 5 sessions, scaled down to 18% of the grade
- A passing threshold of 17.5/35 points
- Designed to improve your paper reading skills

Lab assignments

- Hands-on experience of using the latest NLP tools in Python
- Prepares you for the project
- The lectures and lab assignments complement each other
- You only need to solve and submit your assignments – no demonstrations
- Total of 9 points for 3 assignments
- A passing threshold of 4.5/9 points

Project assignments

- Groups of 3 students work together on a focused TAR topic
- Steps:
  1. Study related work
  2. Prepare the data (collect, clean, annotate)
  3. Implement the system
  4. Evaluate the system
  5. Write up a project report
  6. Give a 10-minute presentation + demonstration
- Expected workload: 39 hours per person over 11 weeks
- Score takes into the account the results, project report, and presentation
  - Group score, to be distributed among and by the group members
  - A limit of 55 points per person (bonus points excluded), contributing directly to 55% of the grade
- Project assignments $\Rightarrow$ soon available at the course website

Project report

- In a form of a scientific paper
- Min. 3 pages and max. 4 pages + 1 page for references
- Structure:
  1. Intro: explain the problem, the motivation, and what you did
  2. Related work: short task overview, similar algorithms/systems
  3. Description of your system
  4. Evaluation of your system
  5. Conclusion
  6. References
- Preferably in English (but we won't insist ☺)
- To be typeset in LaTeX (we provide a template)
- Reviewed by teaching staff + TAs

Project report

- Published on-line in the now renowned "TAR Course Project Report" series (with your consent)
  - Take a look at TAR 2021 PR, TAR 2020 PR, TAR 2019 PR, TAR 2018 PR, TAR 2017 PR, TAR 2016 PR, and TAR 2014 PR
- We highly encourage you to write your report in English and let us publish it on-line
  $\Rightarrow$ gives great EXPOSURE to your work
  $\Rightarrow$ future employers will FIGHT over you! ☺

Grade breakdown

	Continuous		Exam	
	Threshold	% grade	Threshold	% grade
Theory	50%	18%		
Labs	50%	9%		9%
Paper reading	50%	18%		
Project		55%		55%
Written exam			50%	36%

Grading

"I would have had higher grades, but I donated them to charity."

Excellent (5)	**89**
Very good (4)	**76**
Good (3)	**63**
Acceptable (2)	**50**

Project work schedule

- March 4: Topics announced
- March 4–13: Bidding for topics (in teams)
- March 14: Topics assigned
- Apr 4-5: Project progress checkpoint "alpha"
- May 23-24: Project progress checkpoint "beta"
- June 5: Project submission deadline (no extension!)
- June 6–10: Project presentations

Office hours

- Office hours: Tuesdays at 16:00h
- Applications for office hours are to be made via a form on Monday at the latest
- Do not hesitate to pester us, we are here for you, not the other way around

ECTS breakdown

- This course is worth 5 ECTS credits = 130 hours of work
- 46 hrs for the project
- 30 hrs for paper reading
- 26 hrs for the lectures
- 10 hrs for theory pre-reading and study
- 10 hrs for lab sessions
- 8 hrs for project presentations

You now...

1. Know what this course is about and be glad that you've enrolled it
2. Know what are the topics that we want to cover
3. Hopefully became intrigued by some of them
4. Know what you need to do to earn your credits