

Temporal Information Retrieval

Wintersemester 2016/17

<http://www.l3s.de/~anand/tir>

Introduction and Course Overview

Who are we ?

- **Instructor:** Dr. Avishek Anand
- **TA :** Jaspreet Singh
- **L3S Research Center**



- Fun with Web stuff - News, Social Media, Wikipedia, Blogs, Archives, Flickr, Amazon reviews
- Analysis, Retrieval, Mining, Learning, Crawling,...
- Cutting edge research
- Building information systems for Web based data

Temporal Collections

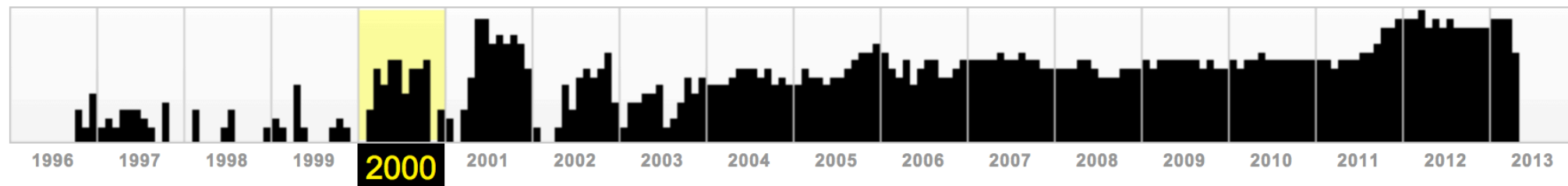


<http://yahoo.com>

Go Wayback!

<http://yahoo.com> has been crawled **14,160 times** going all the way back to [October 17, 1996](#).

A crawl can be a duplicate of the last one. It happens about 25% of the time across 420,000,000 websites. [FAQ](#)



JAN

FEB

MAR

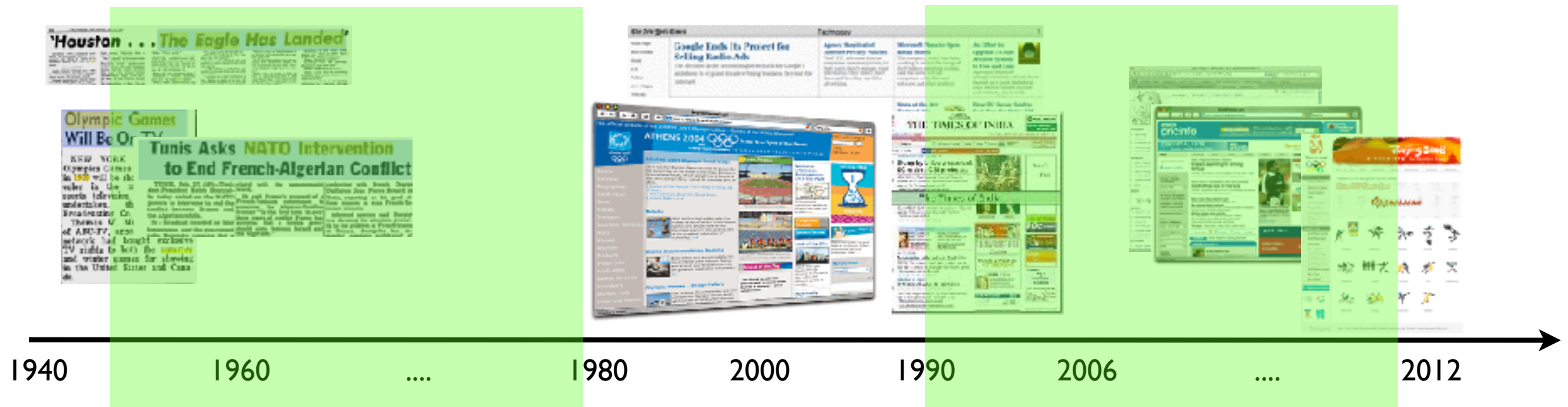
APR



Content increasingly getting digitized..

Content is spanning long time periods and growing

Queries



nato intervention @ 1950-1980

nato intervention @ 1990-2012

- Time travel queries : Keyword queries with time of interest

nato intervention @ 25/12/2001 **nato intervention @ 1950-1980**

Nato Intervention Ashed In French ... Reading Eagle **time points vs time ranges**

- Retrieve documents from the past
- Relevance based on time of interest

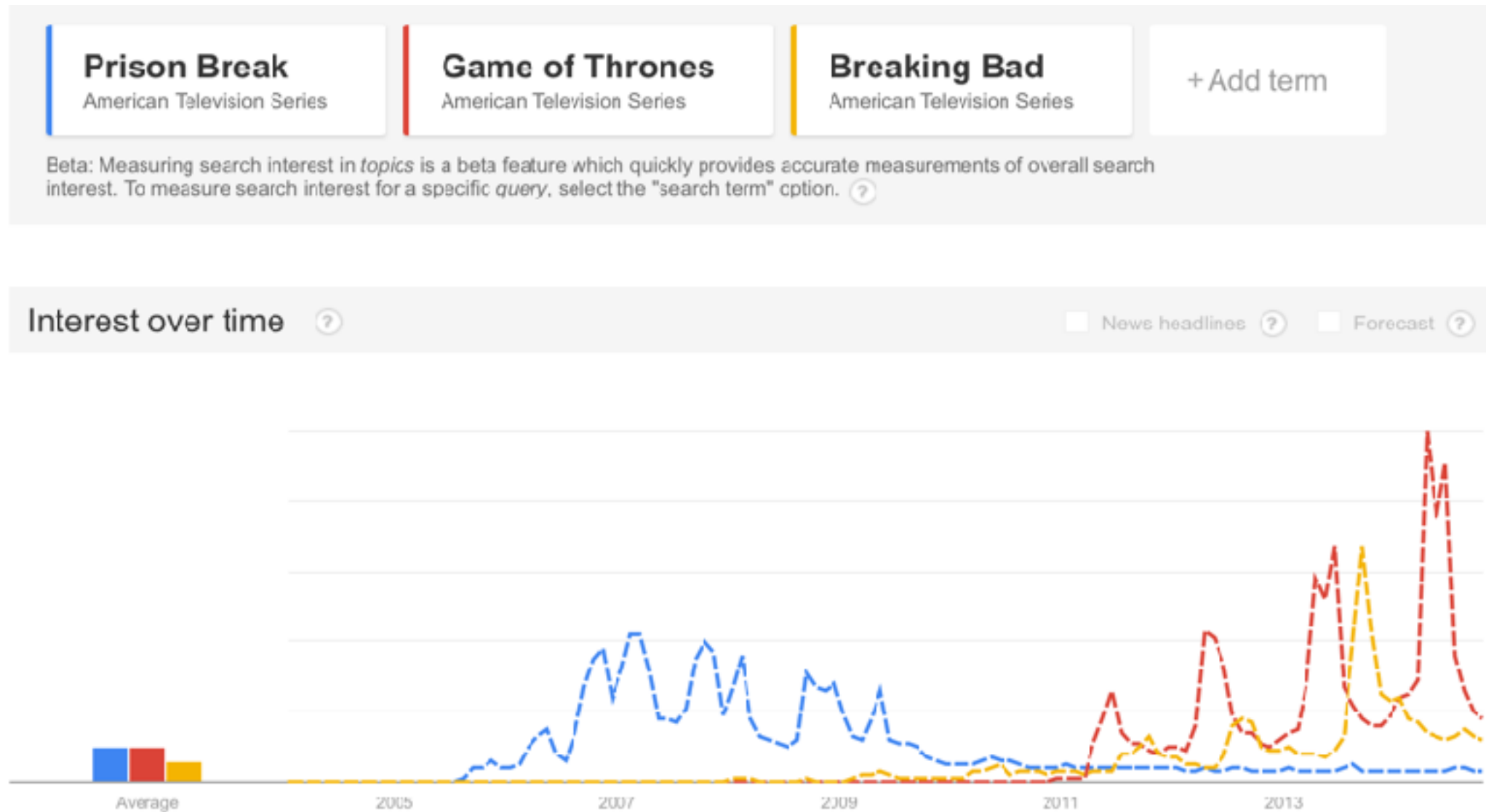
Issues in Temporal IR

- Temporal Collections



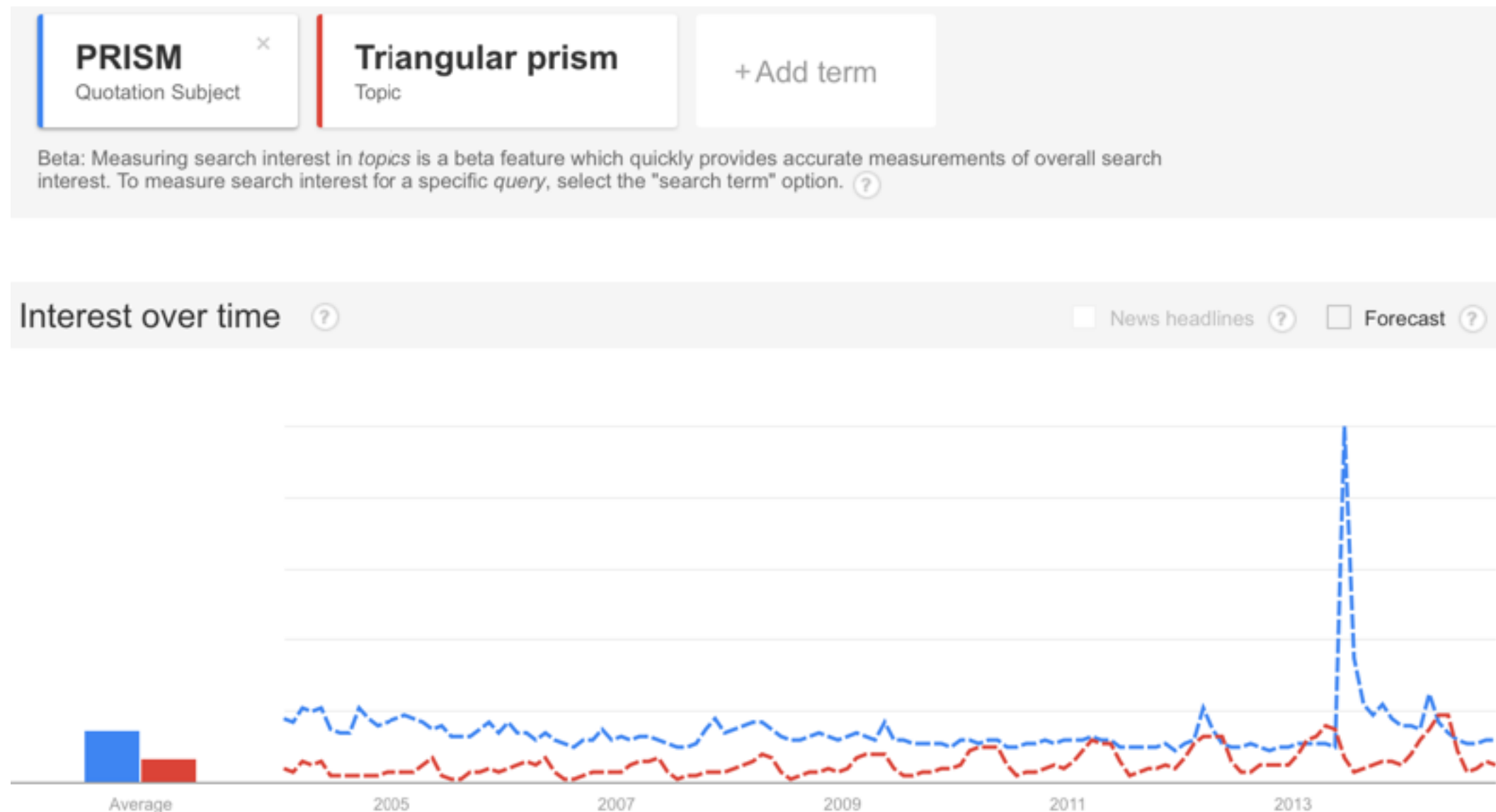
- Web Archives, News Archives, Wikipedia Versions....
- Temporal information needs **nato intervention @ 25/12/2001**
- Historical vs Recency
- Document relevance with temporal relevance
 - Is the doc relevant vs Is the doc relevant at the time of interest ?

Issues in Temporal IR



- Infer popular terms to improve ranking & suggestions

Issues in Temporal IR

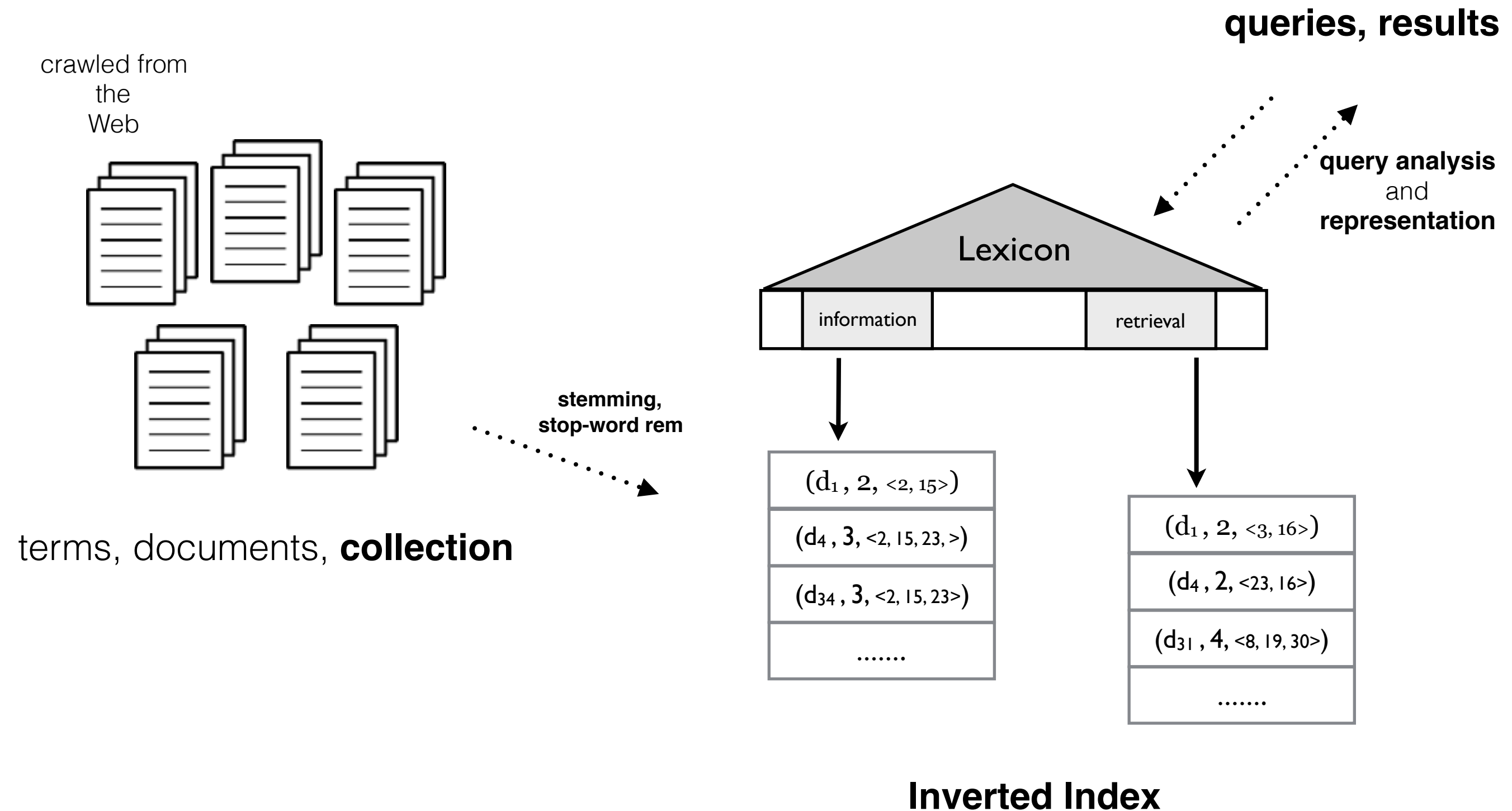


- Infer popular terms to improve ranking & suggestions

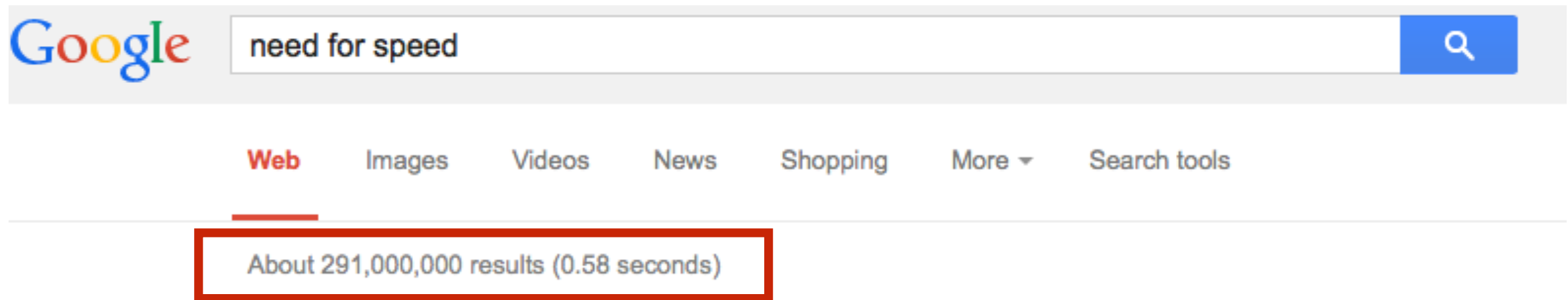
Course Outline

- Foundations of Temporal Analysis
- Indexing for Temporal Retrieval
- Retrieval Models
- Query Analysis
- Word Representations

Simplistic Overview - Web IR System



Need for Speed



- Sub-second retrieval
- Interactive search
- Search as a primitive for further mining and post processing

Indexing

Indexing Temporal Collections



- How do you search these collections efficiently ?
- How do you build and maintain indexes over evolving text collections ?
- What will you learn:
 - **Indexing large text collections,**
 - **Compression and how to handle updates, QP**
 - **Software: Lucene**

Temporal Ranking



- Ranking a large set of results
- Rank results based on relevance to
 - Text in the document
 - Temporal intent
- Temporal Diversity
- Test Collections and Evaluation

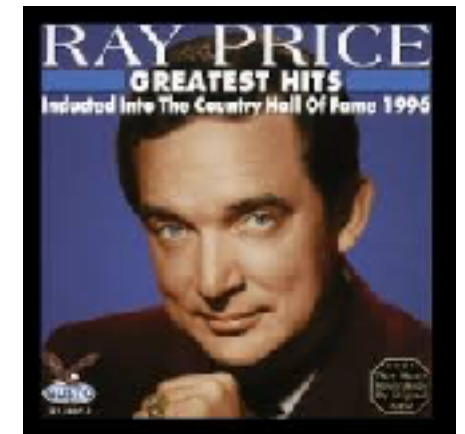
Temporal Ranking



- What will you learn ?
 - Statistical Language Models
 - Support Vector Machines
 - Learning to Rank
 - **Software: RankLib, SVMRank**

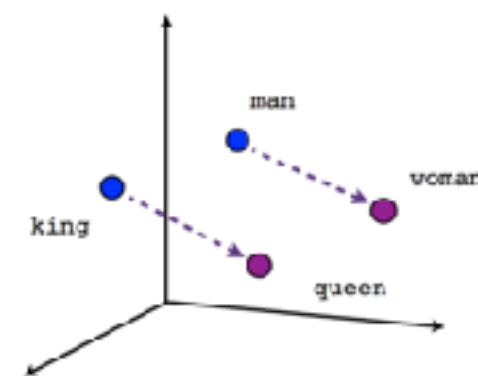
Query Understanding

- What does the user mean by the query ?
- Query : jaguar price

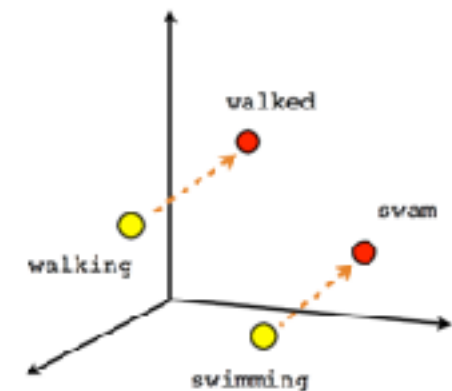


Query Understanding

- Word embeddings
 - Similar words share similar contexts
 - Compositionality
- What will you learn ?
 - Neural Networks (basics)
 - **Software: Word2Vec**



Male-Female



Verb tense

Query Analysis



- Query auto-completion
- Query Suggestion
- Query Expansions

Building a Temporal Search Engine

Rank by

HistDiv

hannover



Results returned: 72 out of 72



Query: hannover - HistDiv

The New York Times 1-6-2005 1
Director of Offshore A.I.G. Unit Was Cited in Australian Inquiry

Business/Financial Desk

... at a unit of the German reinsurer **Hannover** Re, was appointed to the board of Union Excess in June 2000... through several reinsurers, including **Hannover** Re, that made it appear financially stronger than it was. The questionable transactions arranged by **Hannover** Re allowed HIH to overstate its before-tax... Royal Commission. Mr. Elers declined repeated requests for comment. A spokeswoman for

The New York Times 13-9-2005 8
Cost Estimates By Reinsurers Are Climbing

Business/Financial Desk

... **Hannover** Re of Germany warned that its projected \$314 million in claims could climb further

The New York Times 23-6-1987 2
MUSIC: GERMAN CHOIR

Cultural Desk

...LEAD: THE Knabenchor **Hannover** -which sang a program from the German Baroque at St. Bartholomew's... **Hannover** - which sang a program from the German Baroque at St. Bartholomew's Church Friday night... **Hannover** -about 25 boys and slightly fewer young men for the bass and tenor parts - brought to mind... "Der Geist hilft." The Knabenchor **Hannover** impresses

The New York Times 5-7-2005 9
In Germany, the Jobless Work to Keep Their Benefits

Business/Financial Desk

...In January, Benjamin Dath's eight-month search for a job in northwest Germany ended at an unemployment office, which ordered him to paint classrooms at

The New York Times 15-4-2003 3
World Business Briefing | Europe: Germany: Insurer's Profit Rises

Business/Financial Desk

...A leading reinsurance company, **Hannover** Re, said it earned 267 million euros (\$287 million) last year, up from 11 million euros last year but short of its target of 300 million euros. **Hannover** said write-downs on the value of its capital investments hurt its 2002 results. The year-earlier income was held down by the Sept. 11 terrorist attacks in the United States. In its outlook for 2003, **Hannover** said

The New York Times 26-7-2005 10
2 Connected To Reinsurer May Face U.S. Action

Business/Financial Desk

... included the American Reinsurance Company, **Hannover**

The New York Times 22-3-2005 4
A Home With the MetroStars

Sports Desk

... about \$5 million with **Hannover** 96. **BECKHAM STILL BENDING IT** -- David Beckham, the English star who plays

The New York Times 18-8-1990 6
GERMANS ARRESTED ON IRAQI EXPORTS

Foreign Desk

... prosecutor, said the suspects were arrested in raids in Hamburg, **Hannover** and Dreieich, a small... Preussag A.G., a leading West German construction and heavy engineering concern in **Hannover**,

The New York Times 26-11-1980 11

The New York Times 24-2-2004 5
Blossoming Indoors in M.L.S.L.

Sports Desk

...; and in Germany, striker Clint Mathis scored his third goal in four games for **Hannover** 96. **SOCCER REPORT**

The New York Times 30-8-2005 7
Arts, Briefly

The Arts/Cultural Desk

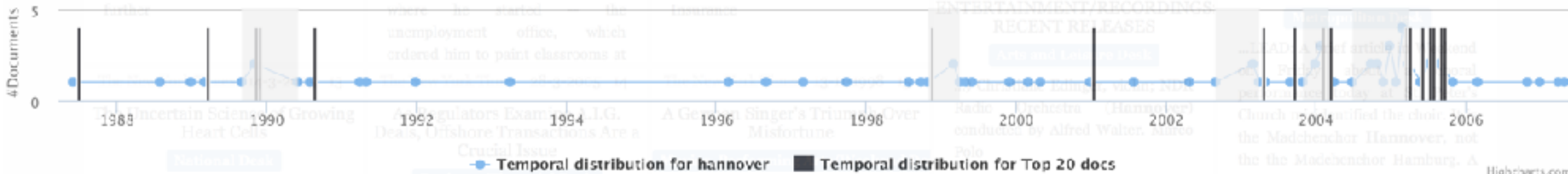
... reported. Mr. Oue, who leads the NDR Symphony Orchestra in **Hannover**, will be replaced by Peter Schneider

The New York Times 5-11-1980 12

4 Documents

Move the mouse over the graph to reveal details. Click and Drag to select time interval.

Corrections



Highcharts.com

Assignments

- **Tutorials:**

- Programming assignments
 - Languages: Python, Java
 - 2 weeks to solve
- Theoretical questions

- ★ **Assignment - 1 [Indexing]**

- ★ **Assignment - 2 [Retrieval Models - LM]**

- ★ **Assignment - 3 [Learning to Rank]**

- ★ **Assignment - 4 [Word Embeddings]**

Final Exam

- **Qualifying for the exam:** 50% in assignments
- **Last lecture:** 24th January
 - Conclusions and Exam revisions
- 30th January : Q&A
- **Final Exam:** Oral Exam
 - **1st Week of February**



Administrivia and Dates

- **Webpage:** <http://www.l3s.de/~anand/tir>
- Assignments, slides and references will be put up every week after the lecture
- **Office hours:** 30 mins after the lecture
- **Recitations:** To improve your foundations, or delve into more details

Demos

- Systems in action
 - Wayback Machine - [link](#)
 - Portugese Web Archive - [link](#)
 - British library - [link](#)
 - L3S HistDiv System - [link](#)