

Temporal Information Retrieval

Wintersemester 2015/16

<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

Digital content is growing continuously...



Temporal Collections

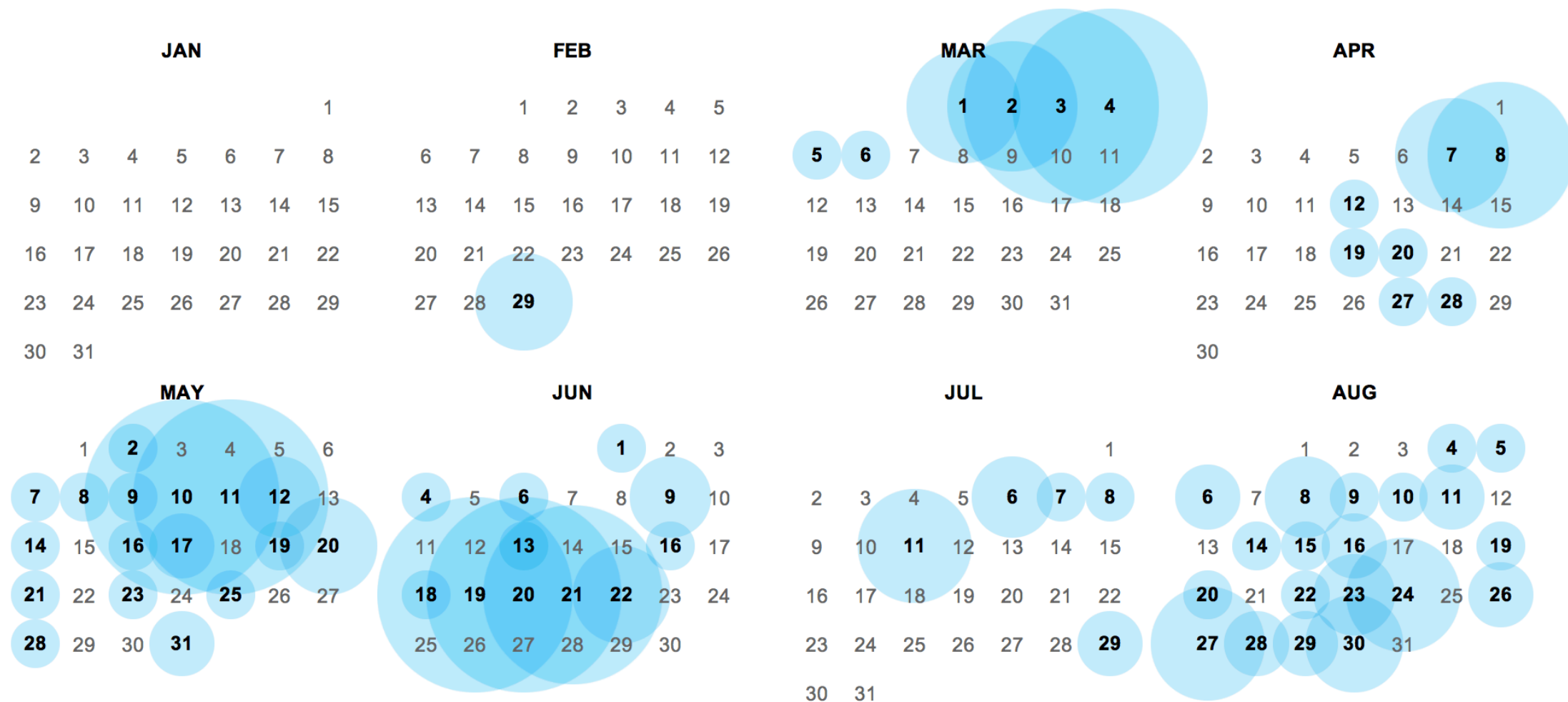
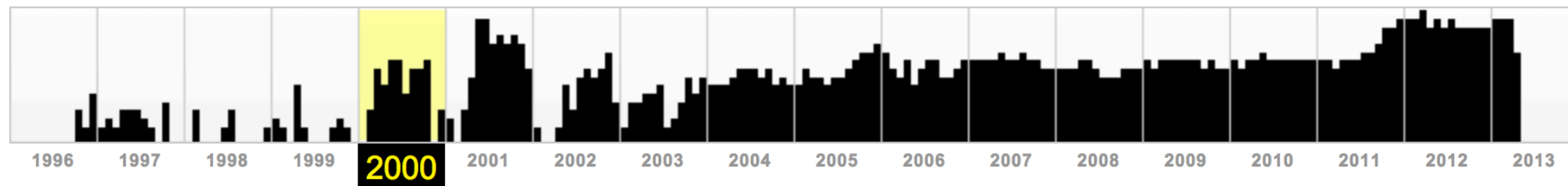


<http://yahoo.com>

Go Wayback!

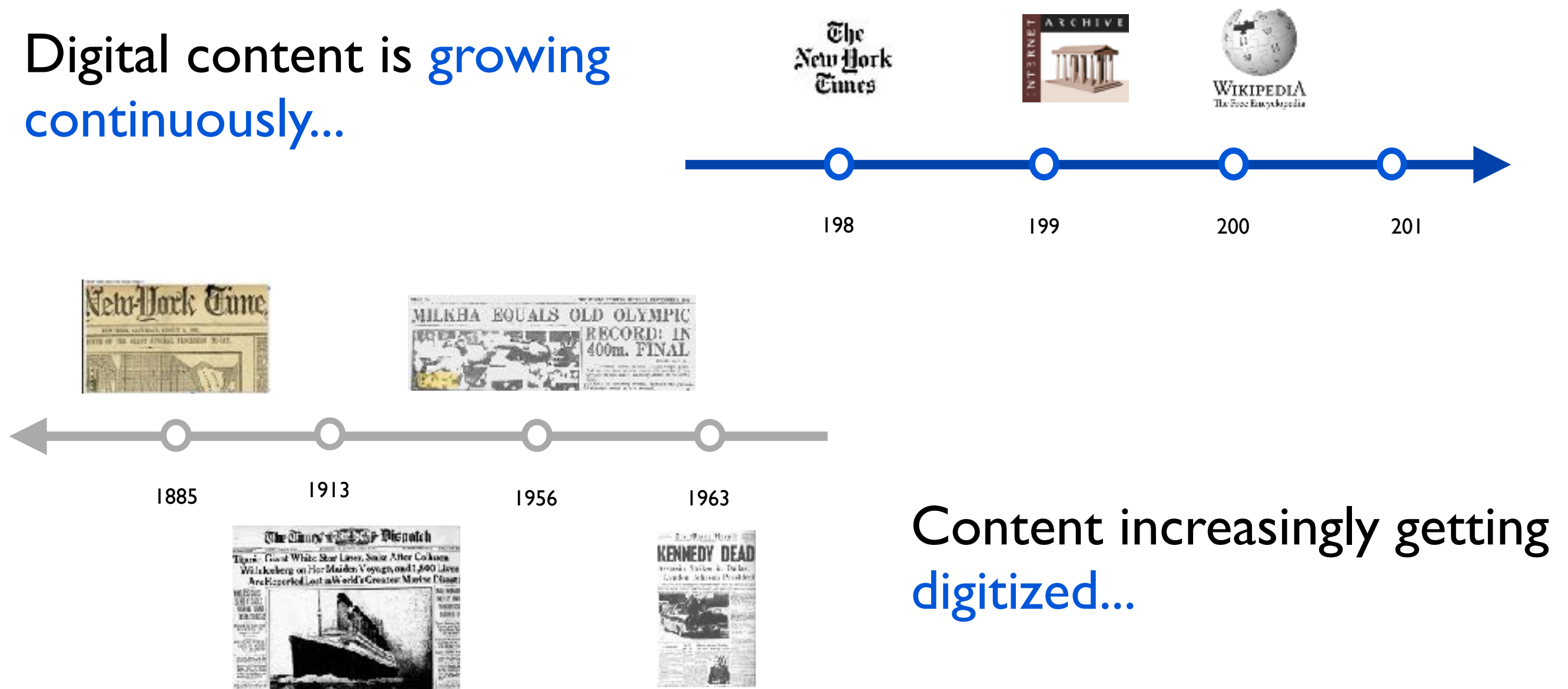
<http://yahoo.com> has been crawled **14,160 times** going all the way back to [October 17, 1996](#).

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



Temporal Collections

Digital content is **growing continuously...**



Temporal Collections

Digital content is **growing** continuously...



Content increasingly getting **digitized...**

Content is spanning long time periods and growing

Queries



1940

1960

....

1980

2000

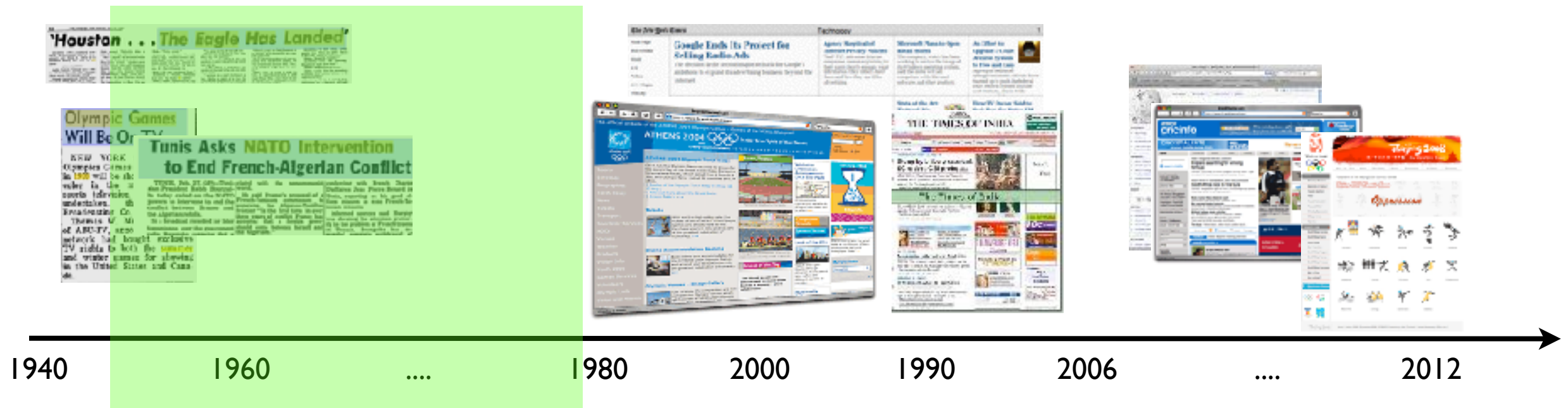
1990

2006

....

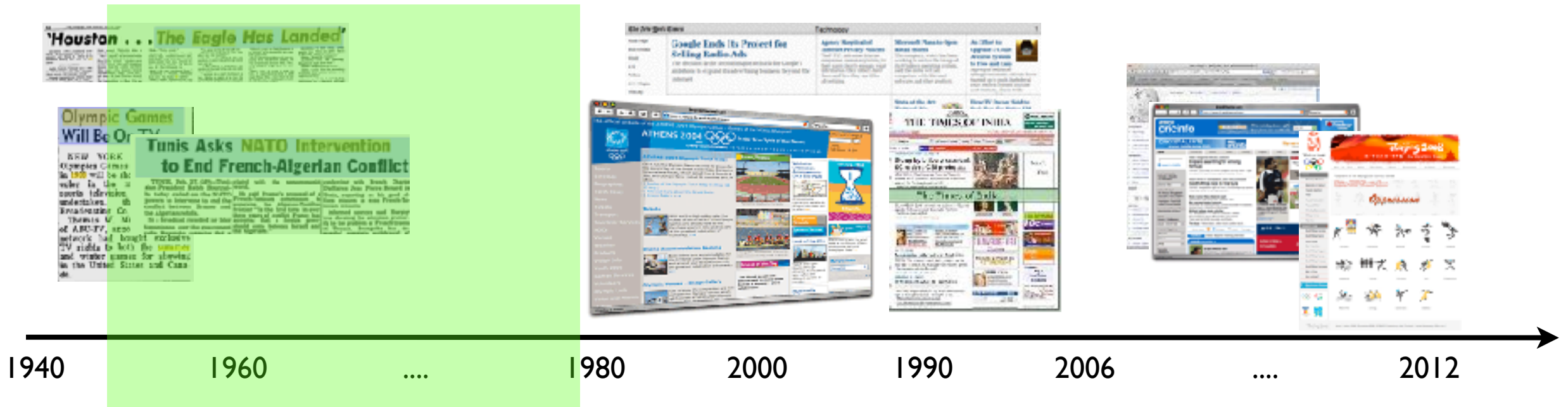
2012

Queries



nato intervention @ 1950-1980

Queries



nato intervention @ 1950-1980

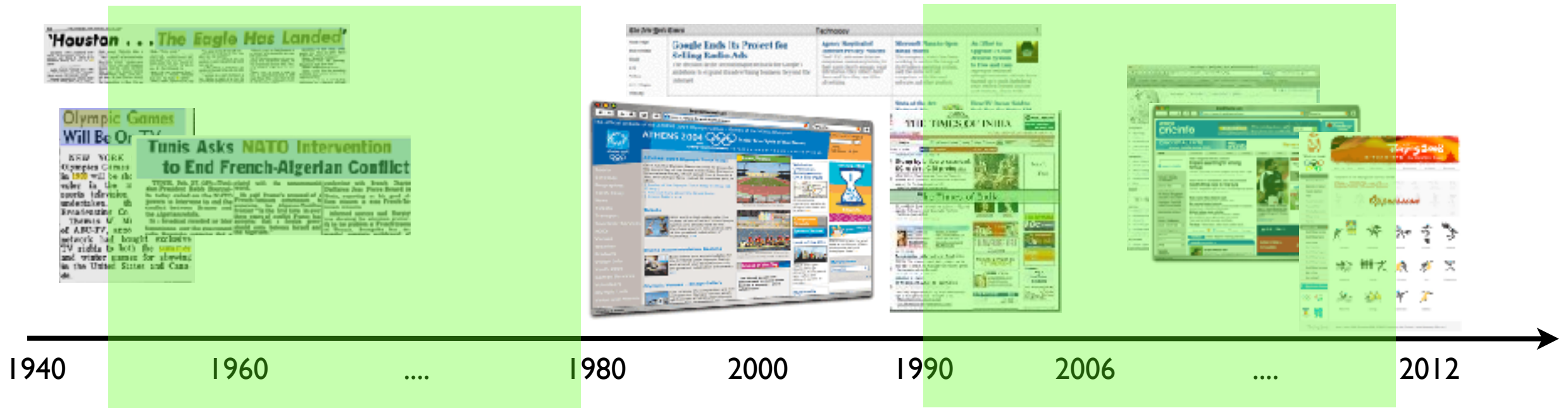
Tunis Asks Nato Intervention To End French-algerian...

Spokane Daily Chronicle - Feb 27, 1958

Habib Bourguiba today called on the **NATO** powers to **intervene** to end the conflict between France and the **Algerian** rebels. In a broadcast recorded for later ...

Nato .Intervention Ashed In French,... Reading Eagle

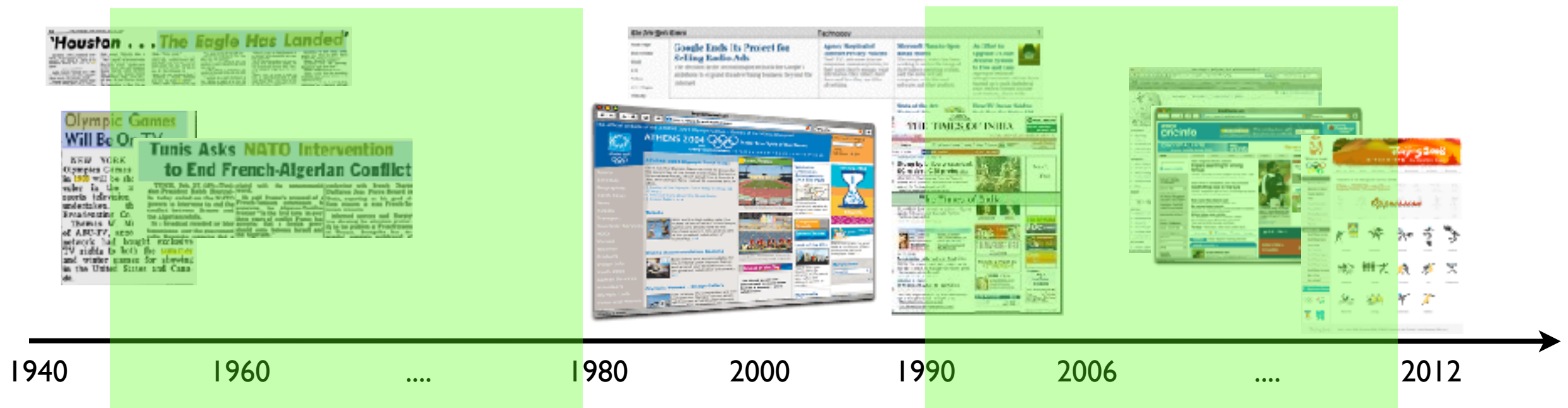
Queries



nato intervention @ 1950-1980

nato intervention @ 1990-2012

Queries



nato intervention @ 1950-1980

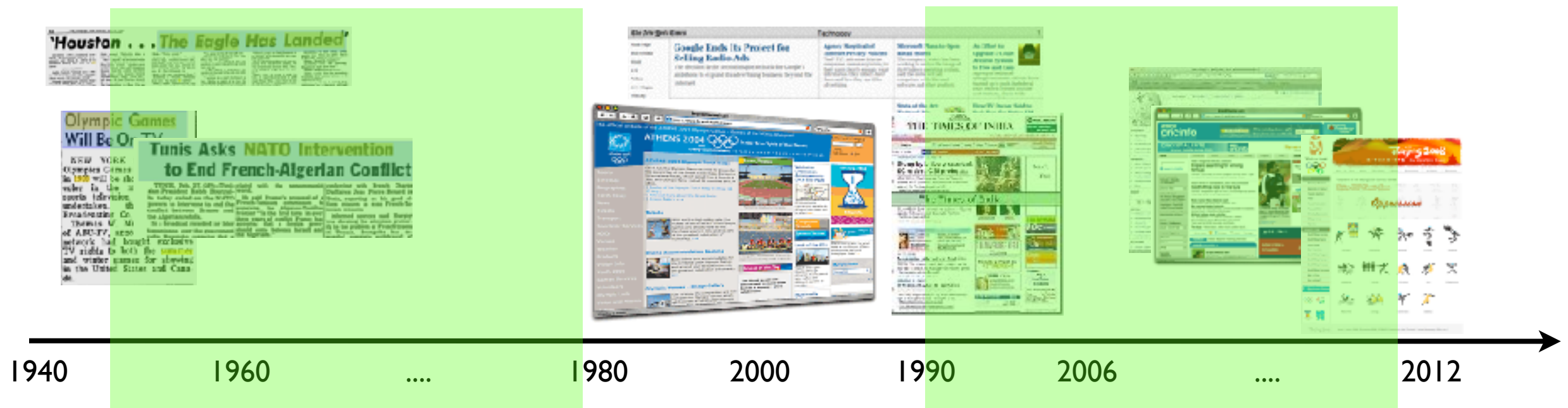
nato intervention @ 1990-2012

- Time-travel queries : Keyword queries with time of interest

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

time points vs time ranges

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

time points vs time ranges

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

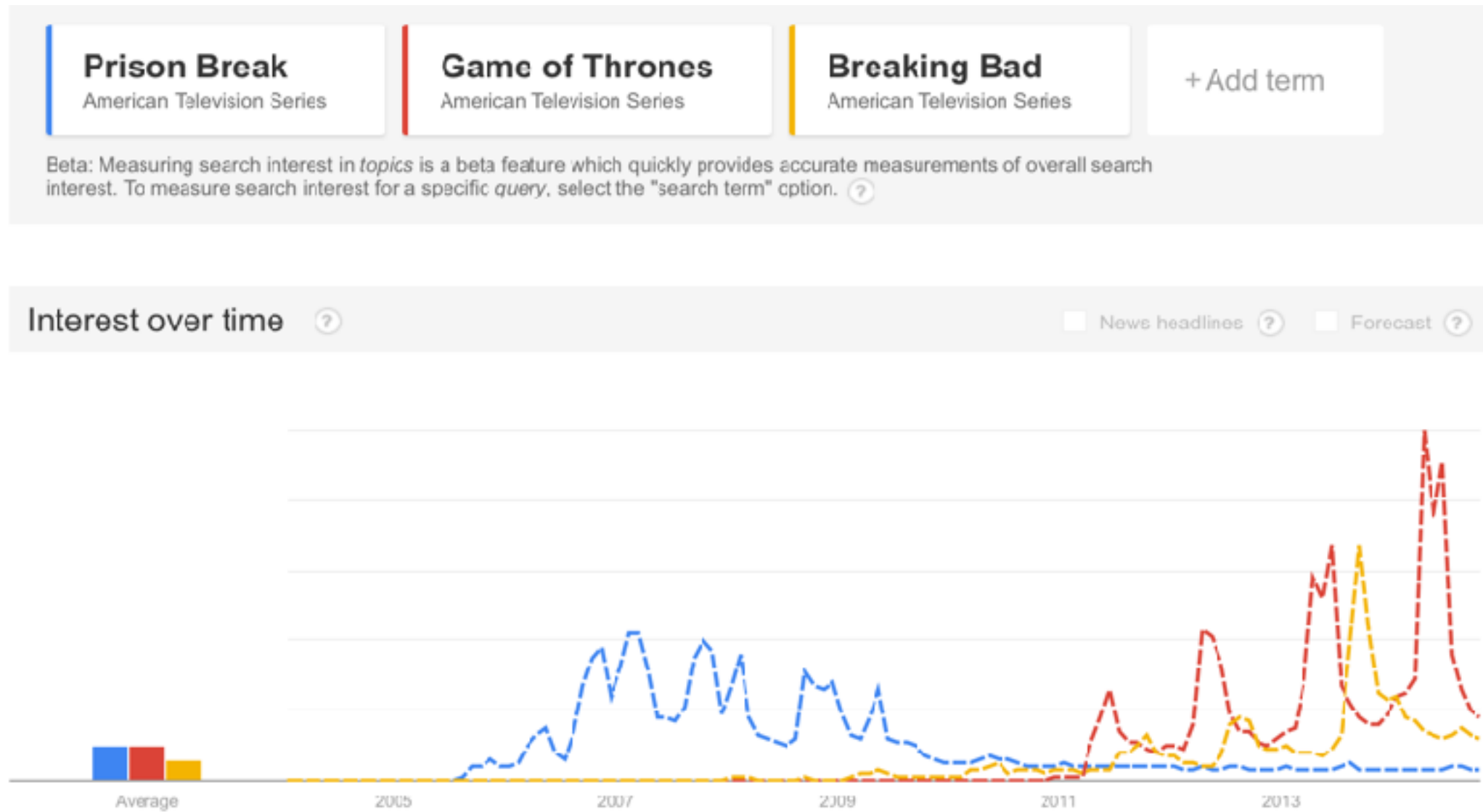
Issues in Temporal IR

- Temporal Collections



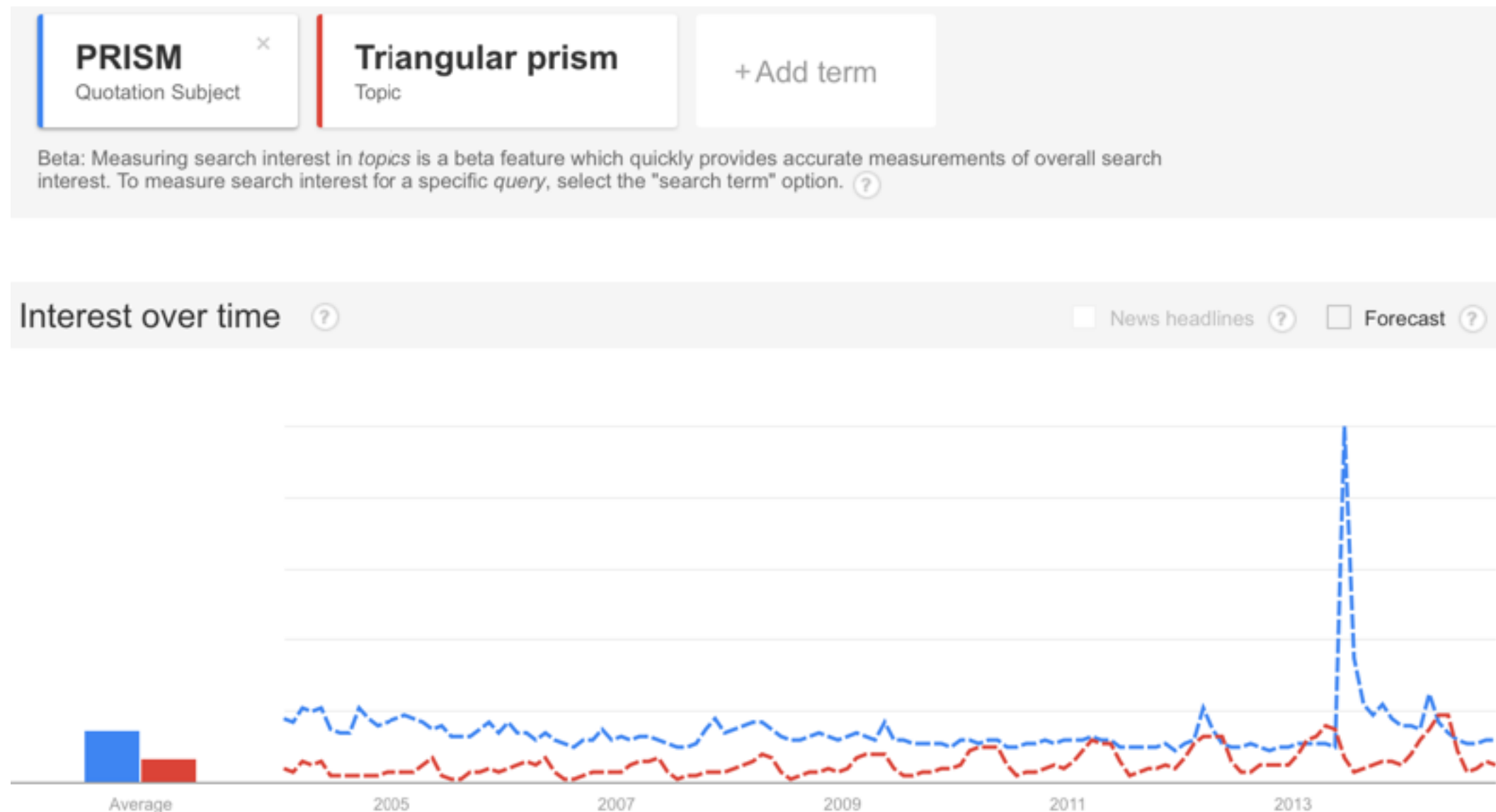
- Web Archives, News Archives, Wikipedia Versions....
- Temporal information needs **nato intervention @ 25/12/2001**
- Historical vs Recency
- Document relevance with temporal relevance
 - Is the doc relevant vs Is the doc relevant at the time of interest ?

Issues in Temporal IR



- Infer popular terms to improve ranking & suggestions

Issues in Temporal IR

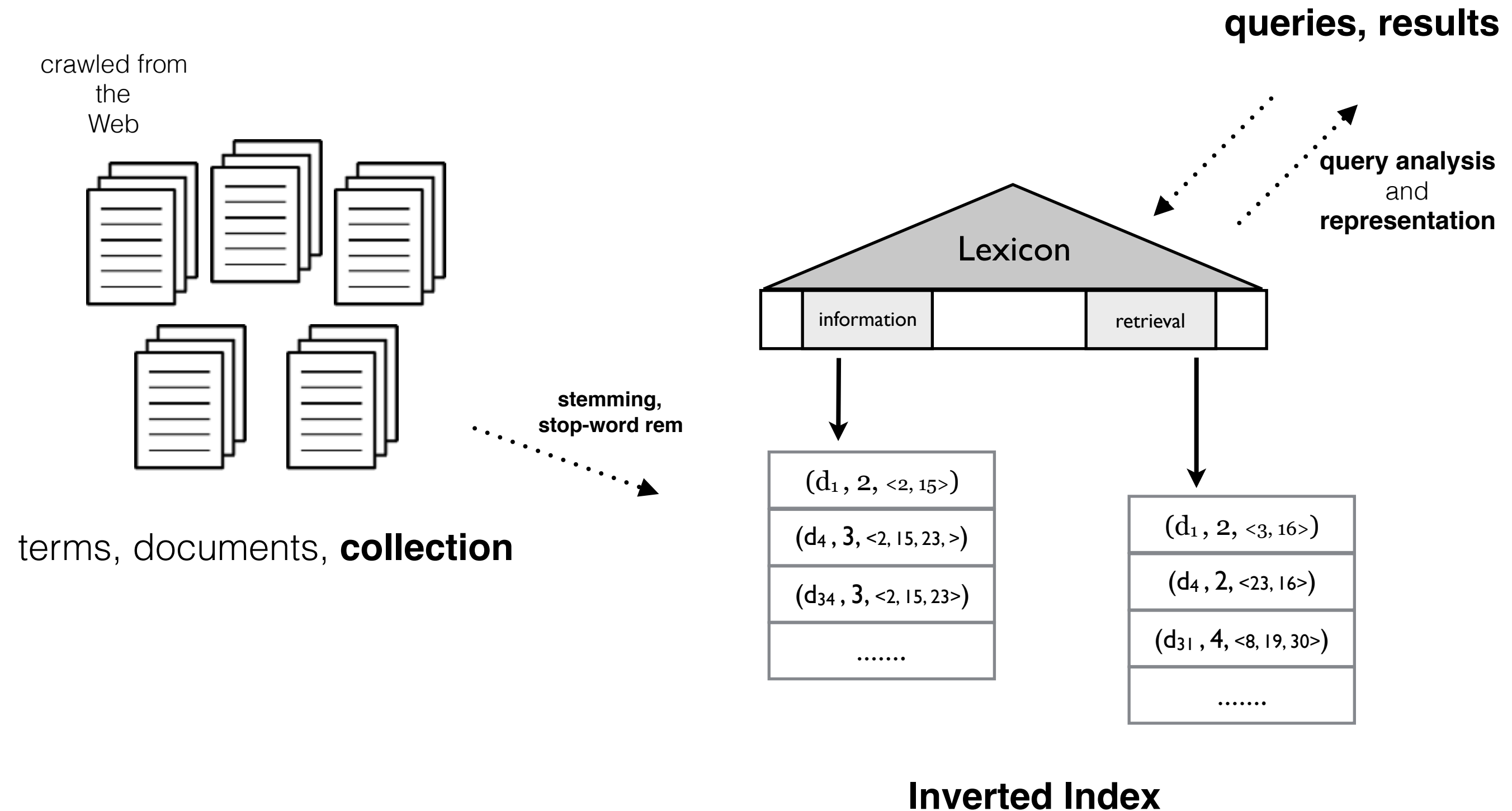


- Infer popular terms to improve ranking & suggestions

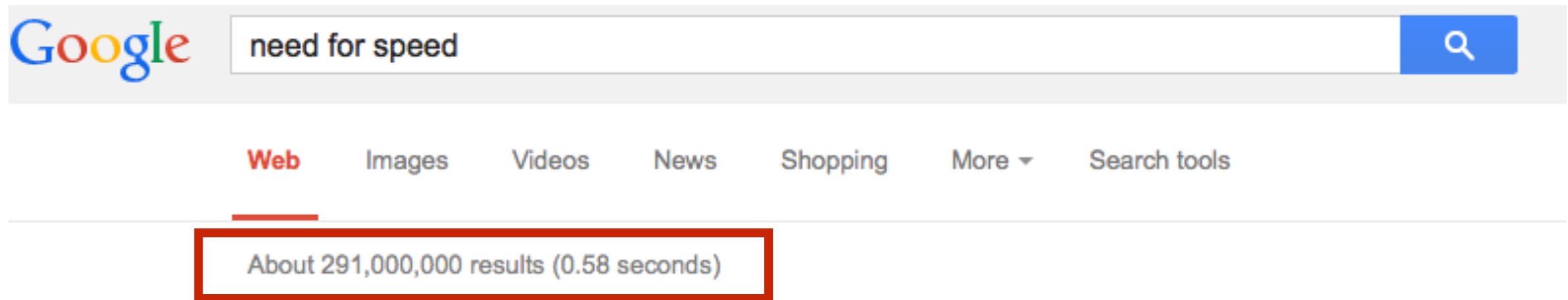
Course Outline

- 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

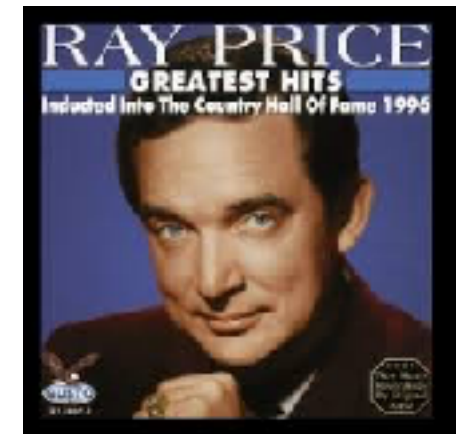
Temporal Ranking



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

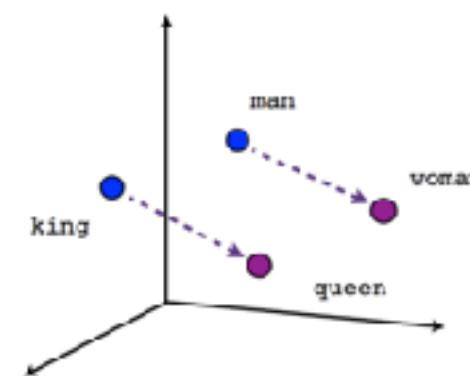
Query Understanding

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

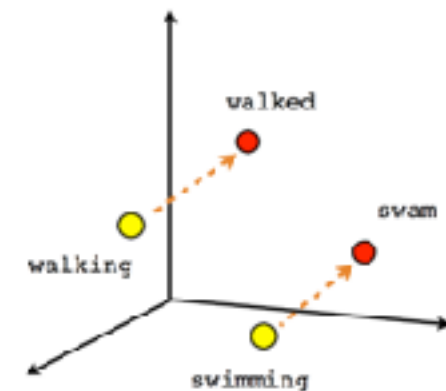


Query Understanding

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



Male-Female



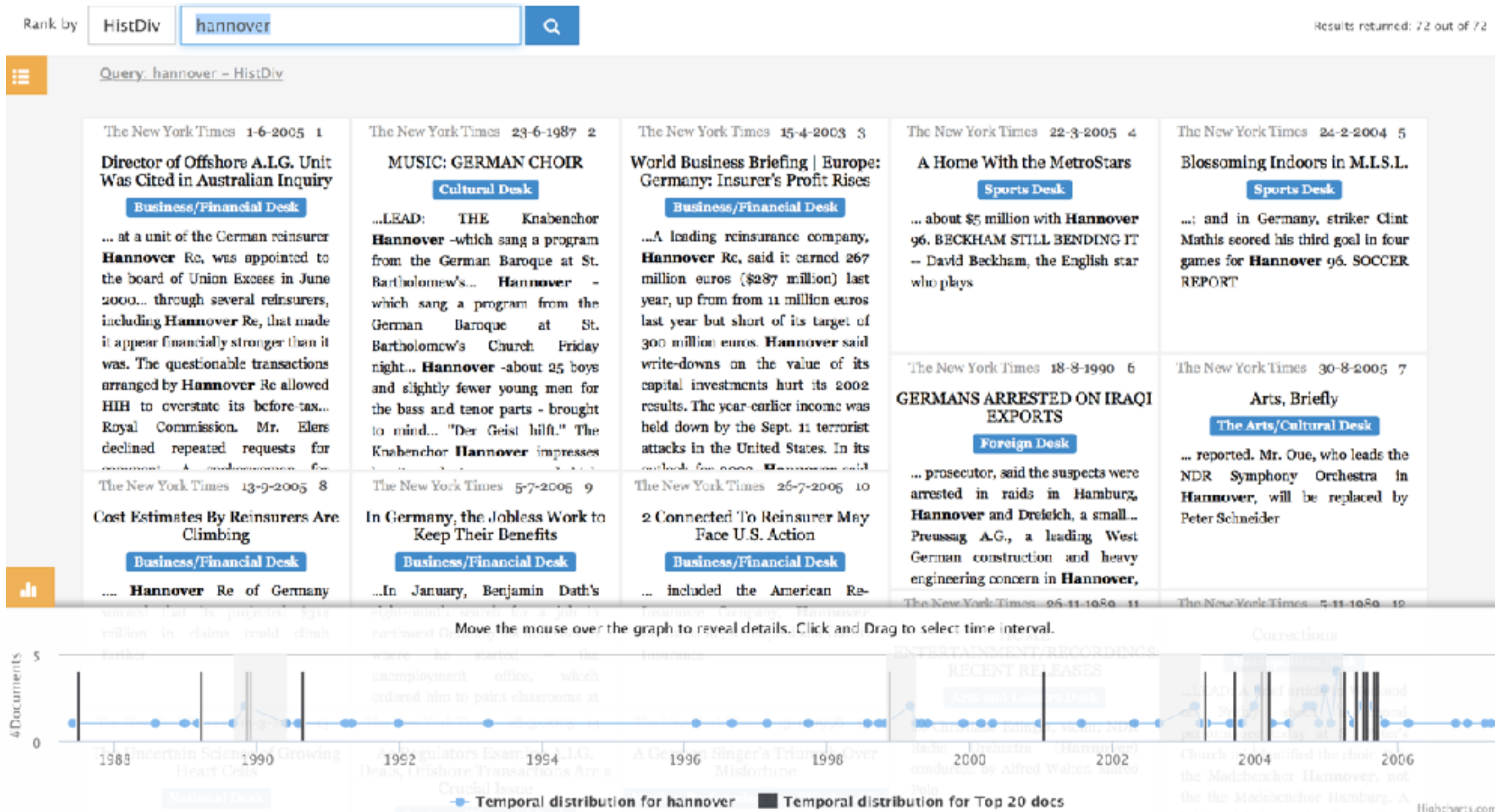
Verb tense

Query Analysis



- Query auto-completion
- Query Suggestion
- Query Expansions

Building a Temporal Search Engine



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

- 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

Demos

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