

# Temporal Information Retrieval

Wintersemester 2014/15

**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

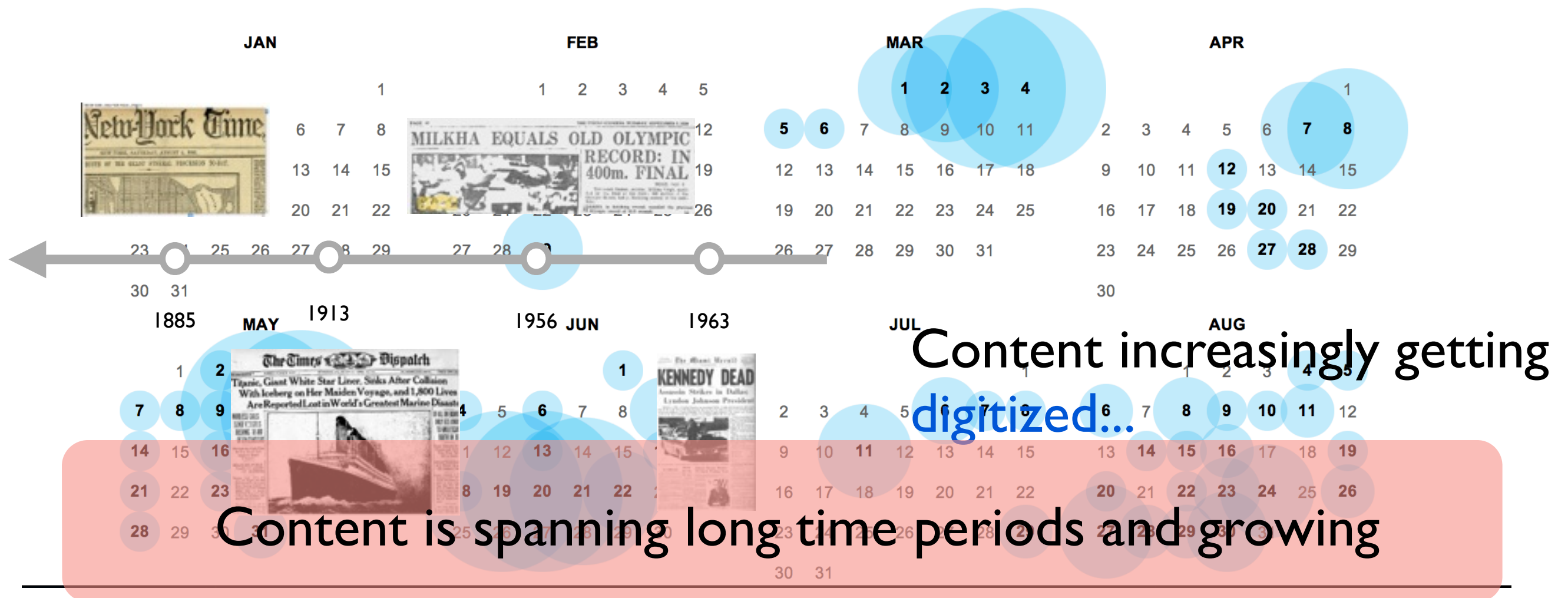
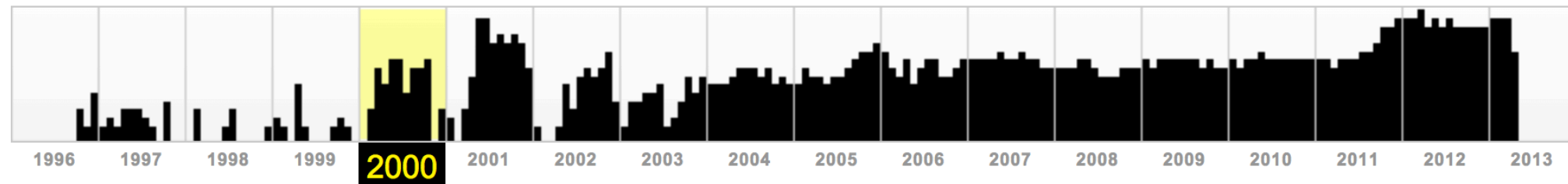


<http://yahoo.com>

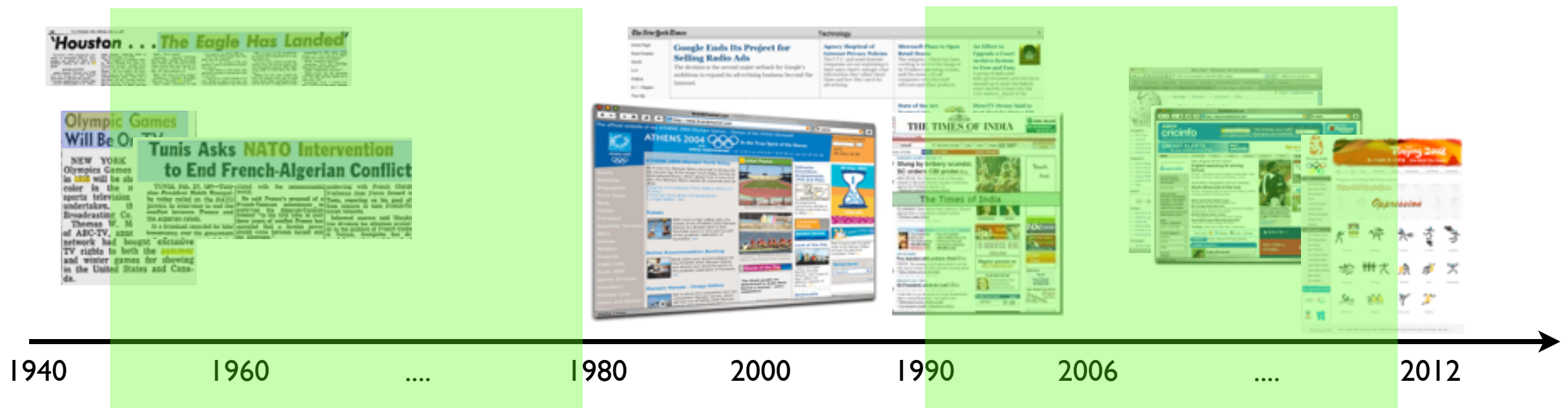
Go Wayback!

<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](#)



# Queries



*nato intervention @ 1950-1980*

*nato intervention @ 1990-2012*

- Time travel queries: Keyword queries with time of interest

*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*

*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*

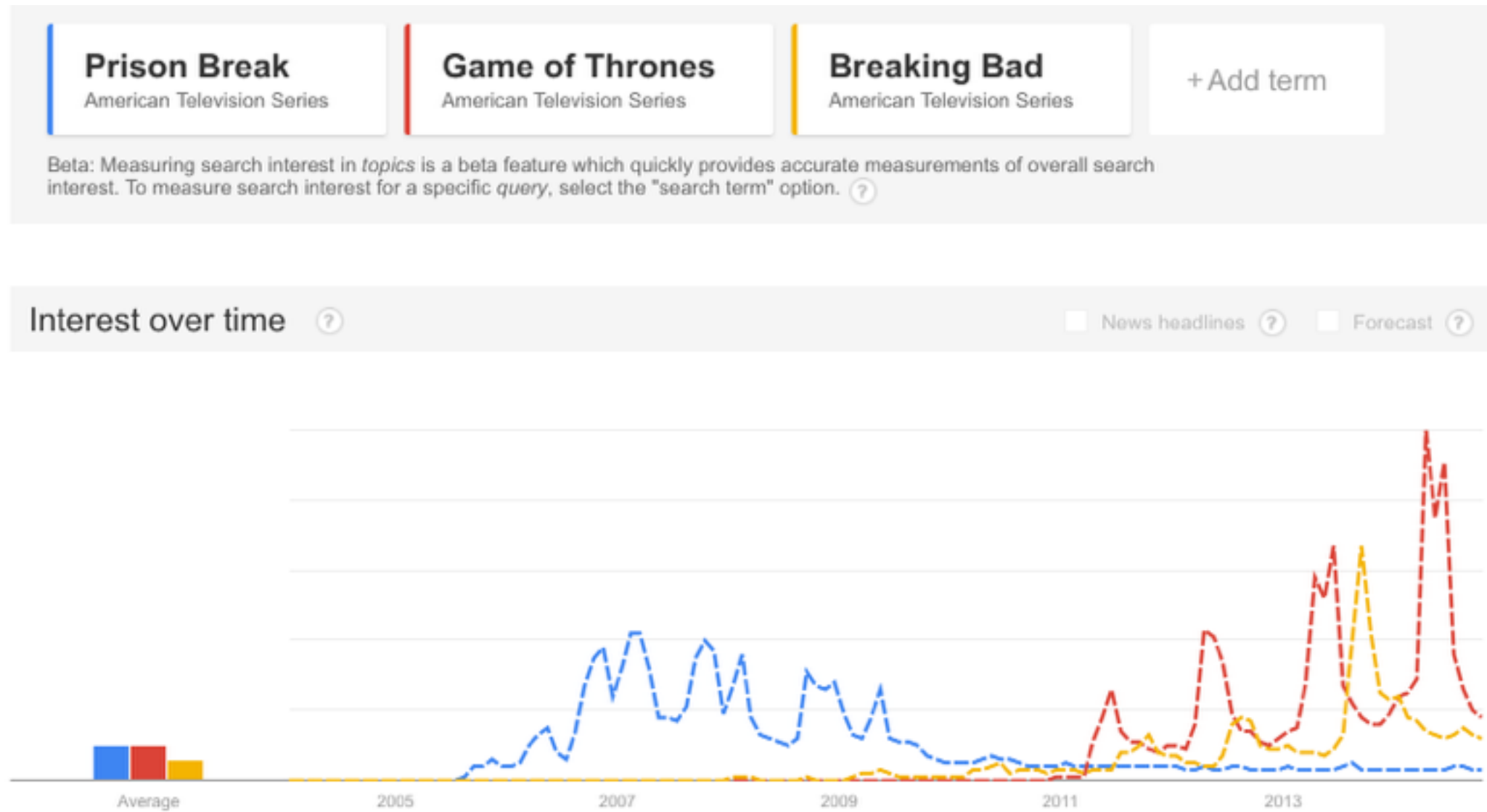
- Historical vs Recency

- 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
- Web Dynamics and Crawling
- Indexing for Temporal Retrieval
- Retrieval Models
- Query Analysis
- Temporal Mining and Applications

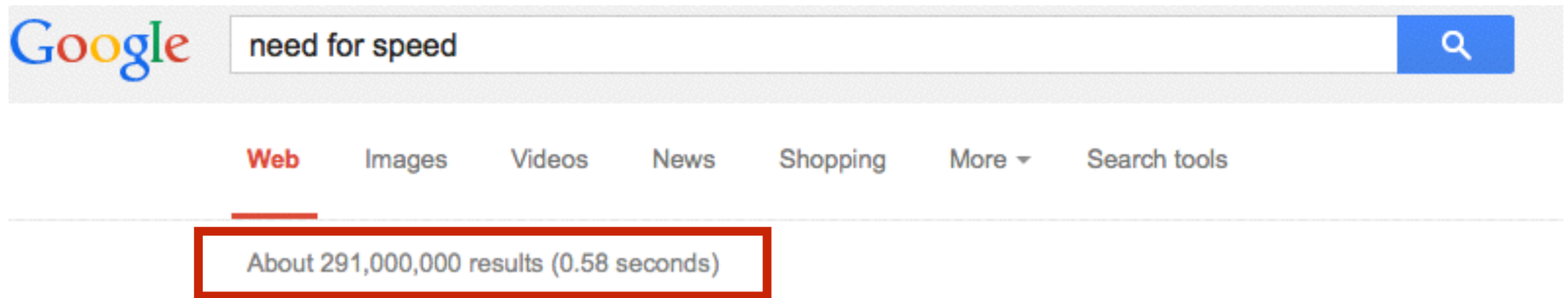


# Catch me if you can

## **Web Dynamics**

- How fast is the Web growing ?
- Change rates, birth and death of Web pages
  - content changes or pages disappear before we crawl them
- How do we make better Crawlers ?
- Learn effective crawling techniques

# 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 ?

# 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

# Query Log Analysis



- Query auto-completion
- Query Suggestion
- Query Expansions

# Temporal Extraction



Document Creation Time: 1998-04-18

Hungarian astronaut Bertalan Farkas is leaving for the United States to start a new career, he said today.

... On May 22, 1995, Farkas was made a brigadier general, and the following year he was appointed military attache

... However, cited by District of Columbia traffic police in December for driving under the influence of ...

- Implicit and explicit temporal expressions



# Building a Temporal Search Engine

LMLM+TMLM+T+DPM2xQuADIA-Select

dennis rodman

Q

BrowseOverviewHelp

The New York Times29-3-19996

Knicks Fall Victim To a Blond Bombshell

Sports Desk

Poor Kurt Thomas. He didn't even see it coming. One minute, he was battling for position under the basket. The next, he was gone -- ejected from the game for throwing Dennis Rodman to the floor. Rodman dusted off his annoying-player manual yesterday. He baited Thomas into losing his senses, positioning a forearm just so, holding his jersey for just a few seconds and using his feet however he could. "He's always kicking," Thomas, the Knicks' power forward, said of Rodman. "That's something he's done for a number of years." On cue. Just when he is about to make his own teammates lose their minds, Rodman somehow finds a way to crawl into the opposing team's psyche and make himself useful again. He scored

The New York Times24-1-199817

Ailing? Angry? Ah, It's Rodman

Sports Desk

Upon hearing that Dennis Rodman had suddenly left the Bulls last night before the Nets game and returned to Chicago, the first thought, naturally, was that he had an emergency appointment with Heidi, his hair stylist. Perhaps to add dye to his latest design, which

The New York Times11-2-20078

Spotlight | N.B.A.; Championship Rings and Nose Rings

Sports Desk

Dennis Rodman has played a lot of roles in his lifetime. He has twice been an N.B.A. All-Star and defensive player of the year. He has also been a bridegroom and a bride, a father and a surrogate son. He has been an author, an actor, a corporate spokesman and a full-time enigma. Rodman, 45, who won five championships

The New York Times28-3-199918

Will Lakers Blow Away All Foes Or Blow Up?

Sports Desk

There's another peroxide blond turning heads at the Great Western Forum, and this one may completely discombobulate the Lakers. A day after he bathed his scalp in peroxide and complained after a loss, Dennis

The New York Times21-10-19937

On the Rebound, Spurs Take Gamble on Rodman

Sports Desk

Dennis Rodman will say to your face what many people say behind his back. "Off the court, I'm a nut," said Rodman, the San Antonio Spurs' talented, but eccentric, forward. But the Spurs, who lost to the Knicks last night, 95-75, in a preseason game at Madison Square Garden, are gambling that Rodman's antics will be overshadowed by what he brings to the team.

The New York Times19-6-19889

Rodman's Defense Stops Big Names

Sports Desk

LEAD: It was just over a year ago that Dennis Rodman made his leap from National Basketball Association obscurity to the headlines. It was just over a year ago that

The New York Times19-6-19889

Rodman's Defense Stops Big Names

Sports Desk






been assigned to guard Larry Bird, said the Celtics' star was "overrated." Rodman said that the only reason Bird had won three successive most valuable player awards was "because

# Teams and Assignments


- Project [Ranking]
- Project [Indexing]
- Project [Temporal Analysis]
- Project [Temporal Extraction]
- Project [Temporal Mining]
- Project [Crawling]

# Grading and Badges



- Final Exam: Oral Exam
- Pick up badges through the course of the lectures
  - Project   
  - Test collection building  
  - Each badge worth 0.3 credit improvement

# Administrivia and Dates

- Webpage: <http://www.l3s.de/~anand/tir14>
- Test collection building: 16th December 
- Assignments, slides and references will be put up every week after the lecture
- Office hours: 30 mins after the lecture
- Final exam : Last lecture day (negotiable)

# Projects

- Not mandatory
- Programming in Java (mostly)
  - Successful completion  
  - Pat on the back (good job)   
  - Make me cry    
- Each project gets a mentor (Jaspreet, Helge, Gerhard, Avishek)
- Report (1 page) + Integration

# Demos

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