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

Digital content is growing continuously...

Content increasingly getting digitized...

Content is spanning long time periods and growing
Queries

Time-travel queries: Keyword queries with time of interest

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

nato intervention @ 1950-1980
nato intervention @ 1990-2012

Time points vs time ranges
Issues in Temporal IR

• Temporal Collections
  • Web Archives, News Archives, Wikipedia Versions….

• Temporal information needs
  • Historical vs Recency

• Document relevance with temporal relevance
  • Is the doc relevant vs Is the doc relevant at the time of interest?

nato intervention @ 25/12/2001
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

Figure 3: (Top) The auto-completion candidates

Figure 4: An example of applying triple exponential smoothing
Temporal Extraction


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...
Building a **Temporal Search Engine**
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](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](#)