

Temporal-Textual Retrieval: Time and Keyword Search in Web Objects



Ali Khodaei¹, Cyrus Shahabi¹, Amir Khodaei²

iCampus/MediaQ iWatch ✓ CT

¹University of Southern California, ²University of California Berkeley

Introduction

- Many Web 2.0 applications are emerging in which data generated by the users have textual (tags) and **temporal** information.
- Users of such applications often need to query the system by providing requirements on time as well as keywords

Lakers Celtics [August 1984- July 1986]

Problem Definition:

Given a number of search keywords, a time interval or timestamp that user is interested in, a temporal-textual web search finds and ranks the most textual and temporally relevant data objects according to query keywords, and the time interval/instance.

Capability

- Inverted File and Interval-Tree Index
- Inverted File Then Interval-Tree Index
- Interval-Tree Then Inverted File Index
- Seamless Tempo-Textual Ranking
 - Uni-Score
 - Dual-Score
- Multiple Timespans
- Decay Functions / Freshness
- Temporal points and regions
- Relative Weights for text and time
- Integration with existing search engines
- Four Approaches:

Motivation

- Web 2.0 applications

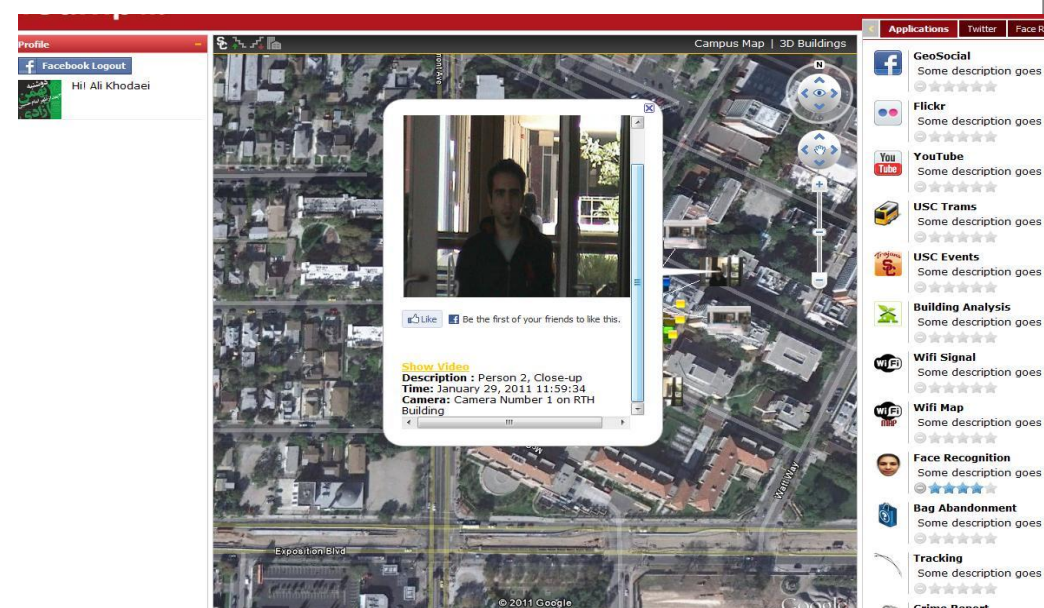
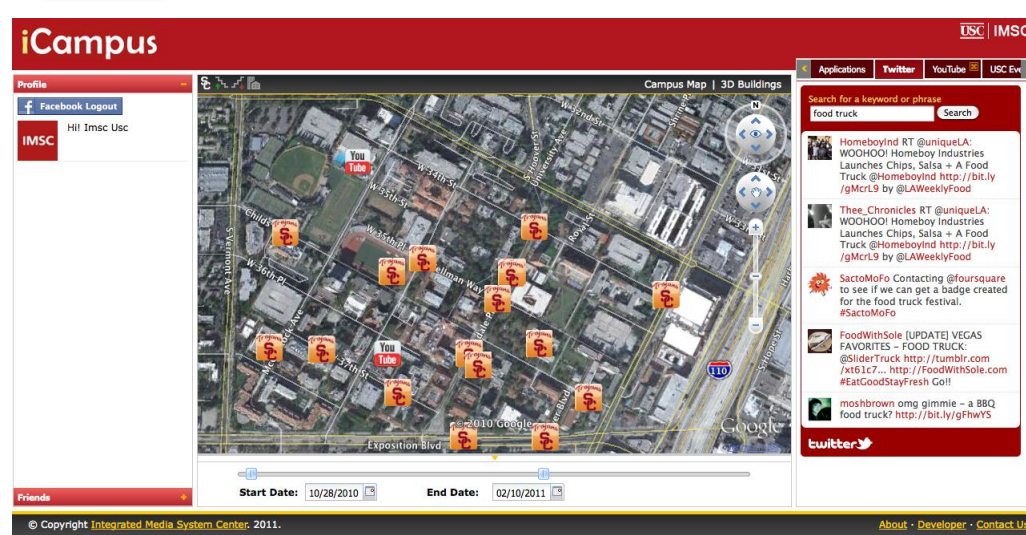


- Web documents

- News, events, biographies, ...

- iWatch

- Numerous heterogeneous surveillance data streams
- Video feeds, suspicious activities, crimes, ...



- iCampus.USC

- Several sources of data with temporal and textual Dimensions

- USC events, USC videos, USC trams, ...

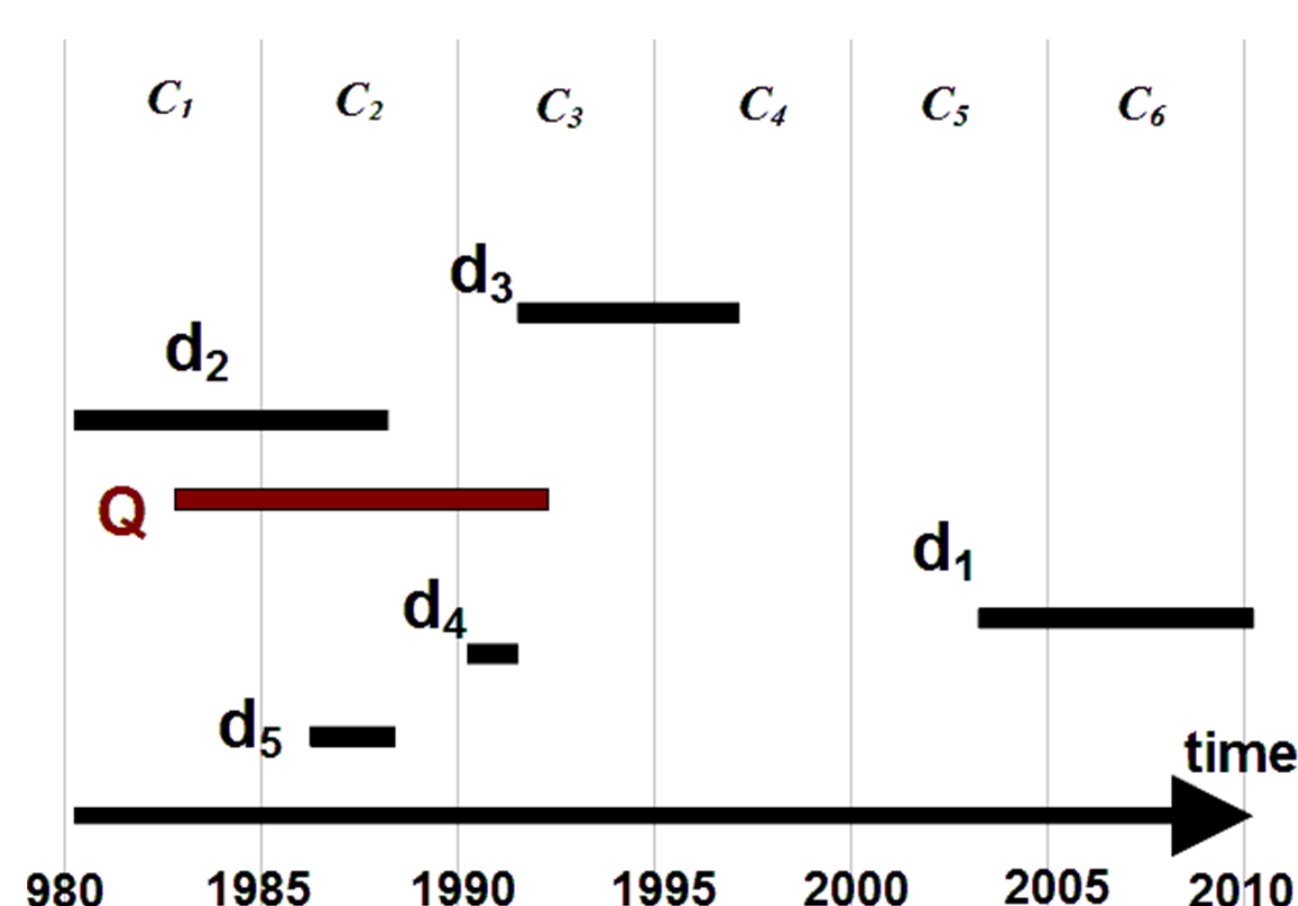
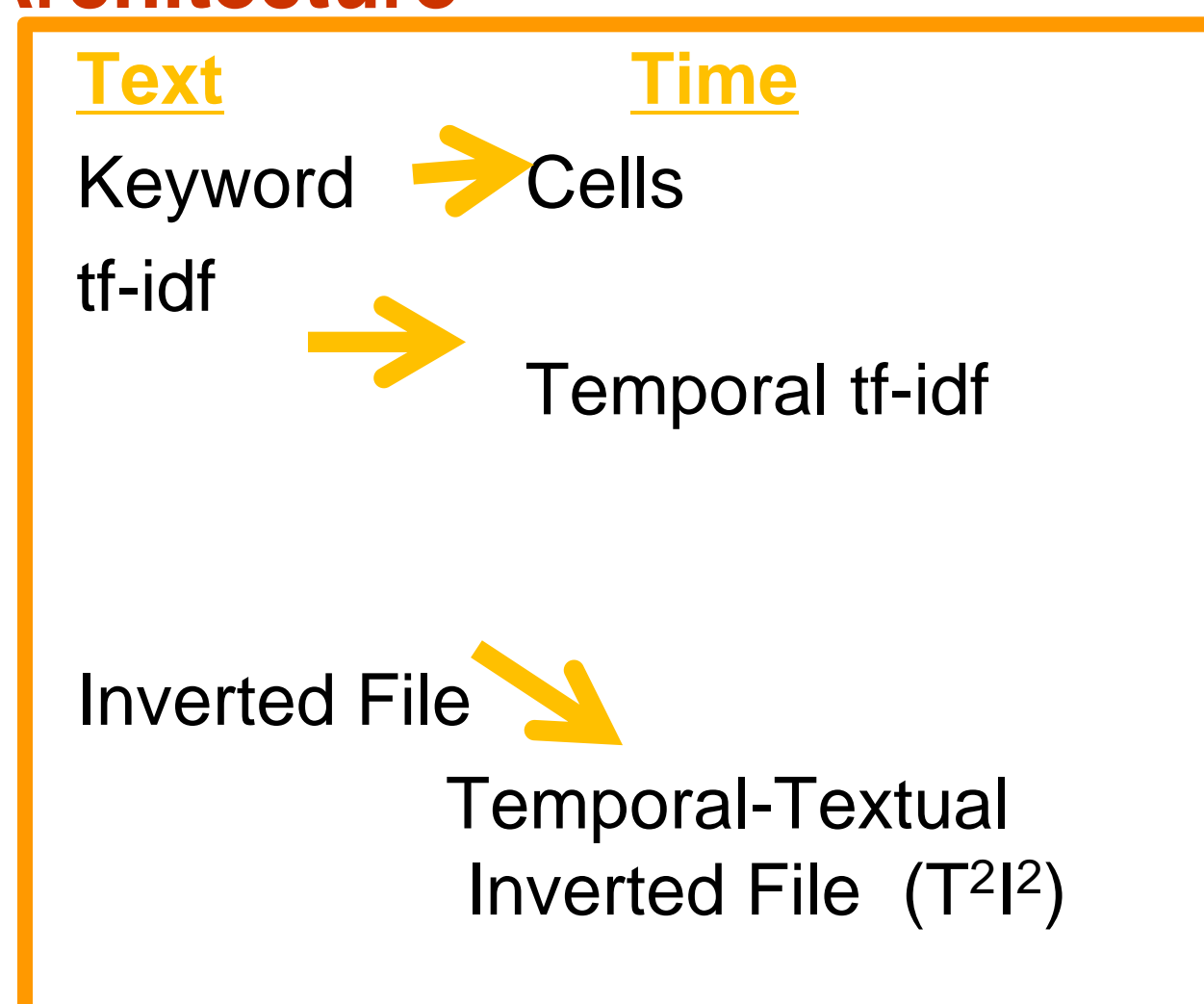
- (1) Uni-score with Inverted document frequency (UI)
Where $ttRel_q(d) = \widehat{S}_{q,d}$ and $w_{q,c} = \ln(1 + \frac{n}{f_c})$
- (2) Uni-score with Direct document frequency (UD)
Where $ttRel_q(d) = \widehat{S}_{q,d}$ and $w_{q,c} = \ln(1 + \frac{f_c}{n})$
- (3) Dual-score with Inverted document frequency (DI)
Where $ttRel_q(d) = \alpha \cdot tmRel_q(d) + (1 - \alpha) \cdot txRel_q(d)$ and $w_{q,c} = \ln(1 + \frac{n}{f_c})$
- (4) Dual-score with Direct document frequency (DD)
Where $ttRel_q(d) = \alpha \cdot tmRel_q(d) + (1 - \alpha) \cdot txRel_q(d)$ and $w_{q,c} = \ln(1 + \frac{f_c}{n})$

System Architecture

- Using same intuitions and concepts used in regular textual searches

- we define new concepts and parameters for temporal data.

- We represent time which is coherent and continuous in nature, as disjunct and set-oriented units of data – similar to the textual keywords



Related Research

- No work on ranking and indexing of temporal and textual features of objects combined
- Textual-temporal index structures
 - Individual index structures / Hybrid index structures
 - Most approaches do not consider the temporal information in the documents' content for the relevance ranking and retrieval .

Conclusion and Future Work

- Add social dimension
- How to Represent, index and rank data with social features
- How to combine it with textual-temporal relevance ranking and indexing