Body:

Motivation

Temporal web dynamics and how they impact upon various components of search systems have received a large share of attention in the last decade. In particular, the study of relevance in information retrieval can now be framed within the so-called time-aware IR approaches, which explains how user behavior, document content and scale vary with time, and how we can use them in our favor in order to improve retrieval effectiveness. In this project, we propose an open-source library, which is served as a research toolkit for temporal search and its related applications.

Challenges & Highlights

This project addresses search challenges centered on the following questions: What are temporal web dynamics? Why do they occur? When and How to leverage temporal information throughout the search cycle and architecture? In particular, we focus on the general and wide aspects associated to temporal web dynamics, from content and structural changes to variations of user behavior and interactions. We also address particular aspects of temporal information extraction (how to timestamp documents and generate temporal profiles of text). The goal is to develop novel concepts, methods and algorithms by taking into account the evolution of entities and events into account, and focus on efficiency and scalability. Our key finding and research results are as follows:

- Large-scale indexing and content analysis methods for temporal web collections
- Mining temporal anchor texts for a large-scale analysis of entities and event evolution
- Scaling up temporal expression extraction and event recognition using MapReduce
- Large-scale time-series analysis and its application for temporal search
- Building temporal PageRank in a collection of billion web documents

Potential applications & future issues

The proposed methodologies and tools can be served as guidance and future directions for similar endeavors, especially applications in related research areas, e.g., natural language processing (NLP), web archiving, and long-term historical studies, where the time dimension also plays an important role.

Project abstract:

Information retrieval (IR) has been extremely successful during the last 20 years in providing an easy access to information available on the Web. Over time, the growing volumes of digital content in the Web have been observed, which demonstrate both its explosive evolution as well as its unprecedented temporal dynamics. The evolution of the Web has a great impact on search. Thus, there is a need for new models and algorithms for understanding temporal developments on the Web and insights into its content and structure changes in order to keep relevant information accessible and meaningful over time.

Project duration:
01.2014 - 12.2014

Bibsonomy show project publications:
1
Bibsonomy use tabs to list publications:
1

Members:
kanhabua
tunguyen
Project manager: Dr. Nattiya Kanhabua

Project research areas: Web Search

Project type: IAI Internal Project
URL: https://www.i3s.de/projects/iai/timetool/

Research Area: Intelligent Access to Information

Status of the Project: