Management of Big Web Data

Course type: lecture
Extent: two hours per week
Website: MOODLE

Outline: in this course we will review advanced topics in Web Data Management. The management of online data involves, among others, storage and effective processing of large volumes of data (as done, e.g., in search engines); accounting for data linkage by explicit hyperlinks or semantic connections; and in the extraction of structured knowledge from raw, semi-structured or unstructured data. We will introduce about prominent tools and principles in these areas.

Topics:
1. Representation of online data
   - HTML, XML, Xpath
2. Searching and ranking of web pages
   - Crawling, PageRank, HITS, top-k algorithms
3. Processing of large datasets
   - MapReduce, Spark, query languages, query evaluation
4. Structuring and retrieving data
   - information extraction, ontologies, RDF, SPARQL
Contents: the course will introduce the following main topics.
(1) Representation of Web data
   - HTML, XML, Xpath
(2) Search and ranking of Web pages
   - Crawling, PageRank, HITS, top-k algorithms
(3) Big data frameworks
   - MapReduce, Spark, query languages, query evaluation
(4) Information extraction and semantic Web
   - information extraction, ontologies, RDF, SPARQL

ו. ת CrossRefים:
דרישות קדם: מערכות מסדי נתונים (89-851) או מסדי נתונים (89-281) או במקביל.
有用的 / דרישות / מטלות: בסיום מסכת, תרגילים במקביל המשגרטר.
မ. လေသာ့စ်ဟွေ့: 85% တပ်ဆင်, 15% တရားလိုင်

Requirements:

Prerequisites: first year courses, database systems (89-851) or databases (89-281) or in parallel

Assignments and other duties: final exam, programming assignments and exercises

Grade: 85% final exam, 15% assignments

Bibliography:

