DEKDIV: A Linked-Data-Driven Web Portal for Learning Analytics
Data Enrichment, Interactive Visualization, and Knowledge Discovery

Yingjie Hu, Grant McKenzie, Jiue-An Yang, Song Gao, Amin Abdalla, and Krzysztof Janowicz

The STKO Lab
Department of Geography
University of California Santa Barbara

Mar. 25, 2014
Outline

1 Motivation

2 DEKDIV: A Quick Flyby

3 Concepts and Methods behind the Scene

4 Architecture and Design Principles

5 Demo
The LAK dataset provides a rich amount of information about researchers, papers, and conferences in the field of learning analytics.

This dataset has already been structured in a machine readable format (RDF).
**Question:** How can we design effective tools to gain more insights from the LAK dataset?
Outline

1. Motivation
2. DEKDIV: A Quick Flyby
3. Concepts and Methods behind the Scene
4. Architecture and Design Principles
5. Demo
DEKDIV: A Quick Flyby

DEKDIV is a unified Web portal consists of a number of functional modules for interactive visualization and knowledge discovery.
DEKDIV: A Quick Flyby

Dataset:

- LAK dataset in RDF
- Concepts and topics extracted from papers’ full text
- Geospatial locations for research institutes and conferences
- External citation data from Microsoft Academic Search as well as expertise information from ArnetMiner
DEKDIV: A Quick Flyby

Interactive Visualization:
- Where do the conference participants come from?
- What are the major concepts in a paper?
- Who are the collaborators of a researcher?
- Which researchers and institutes are citing your papers?
- ...
Knowledge Discovery:

- Who are the most active scholars in a conference?
- How similar are two scholars in their research topic space?
- Who can be suitable reviewers for a newly submitted paper?
- Who can be your potential collaborators?
- ...

DEKDIV: A Quick Flyby
Outline

1. Motivation
2. DEKDIV: A Quick Flyby
3. Concepts and Methods behind the Scene
4. Architecture and Design Principles
5. Demo
Scientometrics and Spatiotemporal Scientometrics

- **Scientometrics** refers to the science of analyzing and measuring a discipline’s research topics, authors, publications and so forth.

- In previous works, we have proposed a framework for **spatiotemporal scientometrics** which focuses on discovering the spatiotemporal patterns of a research discipline (Hu et al., ISWC 2013; Gao et al., ACMGIS 2013).

- **DEKDIV** is an **Online scientometrics workbench** for the LAK community.
Latent Dirichlet allocation (LDA)

- **LDA** is an unsupervised, generative probabilistic model used to infer topics from a textual corpus.
- **LDA** represents each topic using a mixture of keywords. Each paper is described as a number of topics with associated probability values.
- **LDA** has also been applied to describe the interested topics for each researcher.
Multidimensional Scaling (MDS)

- **MDS** is a method to visualize the similarity among individual cases in a dataset.
- **MDS** can take multiple attributes of individual cases, and put these cases into N-dimensional space so that the between-object distances can represent the corresponding similarity.
- In this work, **MDS** has been used to represent the similarity between researchers based on their research topics extracted using **LDA**.
An academic graph can be constructed by considering researchers as nodes and co-authorships as edges.

From each node in the graph, we calculate its shortest path distance to other nodes (researchers) using Dijkstra’s Algorithm.

The shortest path distance has been used to represent how far (or how close) that two researchers are connected.

Application: finding potential collaborators for researchers.

\[
potential = \text{sim}(a_1, a_2) \times (1 - 1/d)
\]
Outline

1. Motivation
2. DEKDIV: A Quick Flyby
3. Concepts and Methods behind the Scene
4. Architecture and Design Principles
5. Demo
A Self-contained Modular Design Paradigm

- The Web portal is designed as a general framework, and each function is implemented as a self-contained module.
Functional Scalability

- Users can "plug" or "unplug" functional modules by adding or removing the import statement of the corresponding JavaScript files.
Data Scalability

- A customized script can be designed to synchronize data and ensure the data in the triple store is up to date.
- This strategy has already been used in the Linked Data portal of the Semantic Web Journal (http://semantic-web-journal.com/SWJPortal/).
Usability Design

- **Simple and intuitive user interface**: Four access points (conferences, researchers, papers, and analytics); “follow your nose” data exploration.
- **Configurable layout of modules**: Each module is displayed in a separate window whose position can be configured based on users’ preference.
- **Rich help document**: Each module is accompanied with a help description, and a Youtube tutorial video has also been provided to facilitate the understanding of the system.
Technology Stack

- **Client side**: JQuery, ExtJS, HTML5, CSS, D3, Leaflet, ...
- **Server side**: Java Servlet, Jena API, SPARQL queries, Jena Fuseki, ...
- **Client-server interaction**: AJAX requests and responses
- **Data models/formats**: RDF, GeoJSON/TopoJSON, ...
- **Performance strategies**: dynamic information caching for frequently requested authors and papers, ...
Outline

1. Motivation
2. DEKDIV: A Quick Flyby
3. Concepts and Methods behind the Scene
4. Architecture and Design Principles
5. Demo
DEKDIV...

CONFERENCES

RESEARCHERS

PUBLICATIONS

ANALYTICS

...powered by