# 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

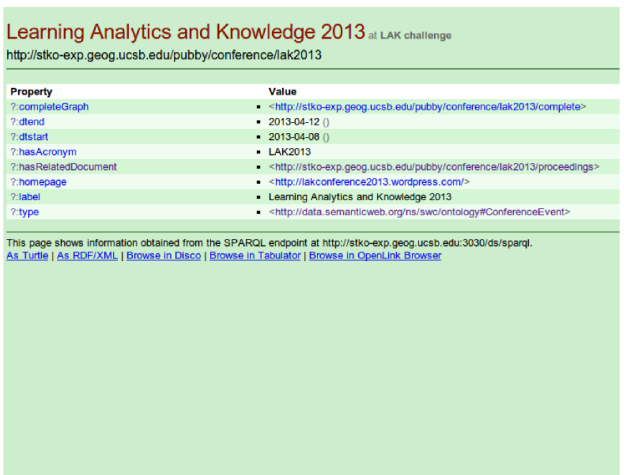
Mar. 25, 2014

# Outline

# Outline

# Motivation

- 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).



Learning Analytics and Knowledge 2013 at LAK challenge

http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013

| Property | Value |
|---|---|
| ?:completeGraph | <http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/complete> |
| ?:dtend | 2013-04-12 () |
| ?:dtstart | 2013-04-08 () |
| ?:hasAcronym | LAK2013 |
| ?:hasRelatedDocument | <http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/proceedings> |
| ?:homepage | <http://lakconference2013.wordpress.com/> |
| ?:label | Learning Analytics and Knowledge 2013 |
| ?:type | <http://data.semanticweb.org/ns/swc/ontology#ConferenceEvent> |

This page shows information obtained from the SPARQL endpoint at http://stko-exp.geog.ucsb.edu:3030/ds/sparql.
As Turtle | As RDF/XML | Browse in Disco | Browse in Tabulator | Browse in OpenLink Browser

proceedings at LAK challenge

http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/proceedings

| Property | Value |
|---|---|
| ?:booktitle | Proceedings of 3nd Conference on Learning Analytics and Knowledge (LAK2013), 8 |
| ?:hasPart | <http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/100><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/101><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/102><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/103><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/104><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/105><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/106><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/107><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/108><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/109><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/110><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/111><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/112><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/113><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/114><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/115><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/116><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/117><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/118><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/119><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/120><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/121><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/122><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/123><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/77><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/78><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/79><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/80><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/81><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/82><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/83><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/84><br><http://stko-exp.geog.ucsb.edu/pubby/conference/lak2013/paper/85> |

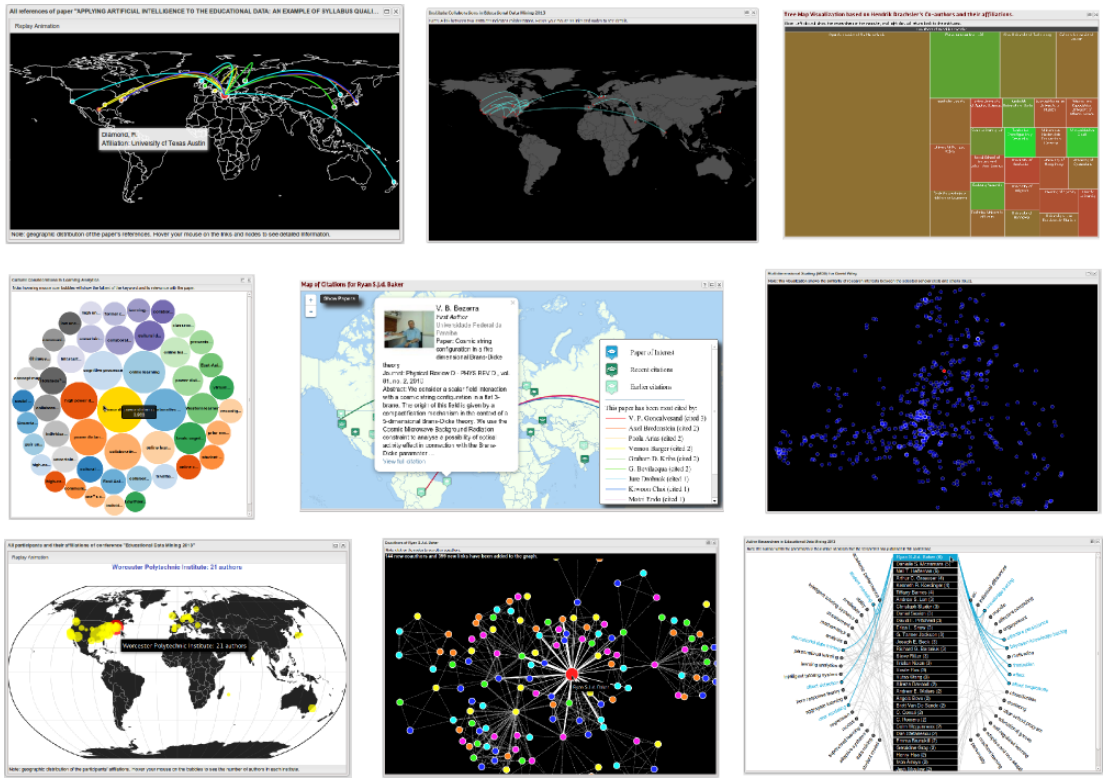
# Motivation

**Question**:
How can we design effective tools to gain more insights from the LAK dataset?

# Outline

# 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:

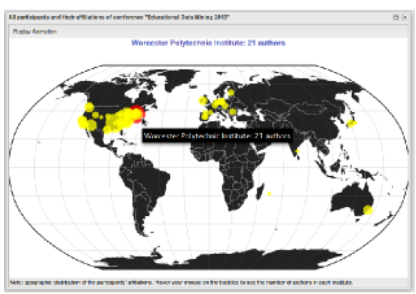
- LAK dataset in RDF
- Concepts and topics extracted from papers' full text
- Geospatial locations for research institutes and conferences
- External citation data from Microsoft Academic Search as well as expertise information from ArnetMiner

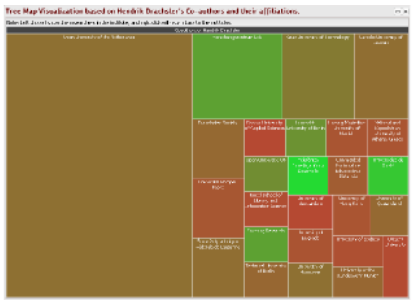



# DEKDIV: A Quick Flyby

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?
- ...

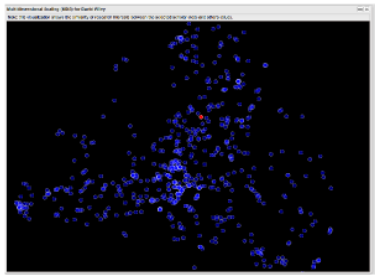
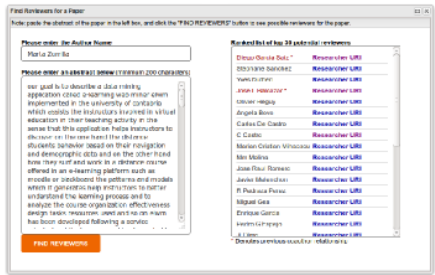



# DEKDIV: A Quick Flyby

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?
- ...

# Outline

# Scientometrics and Spatiotemporal Scientometrics

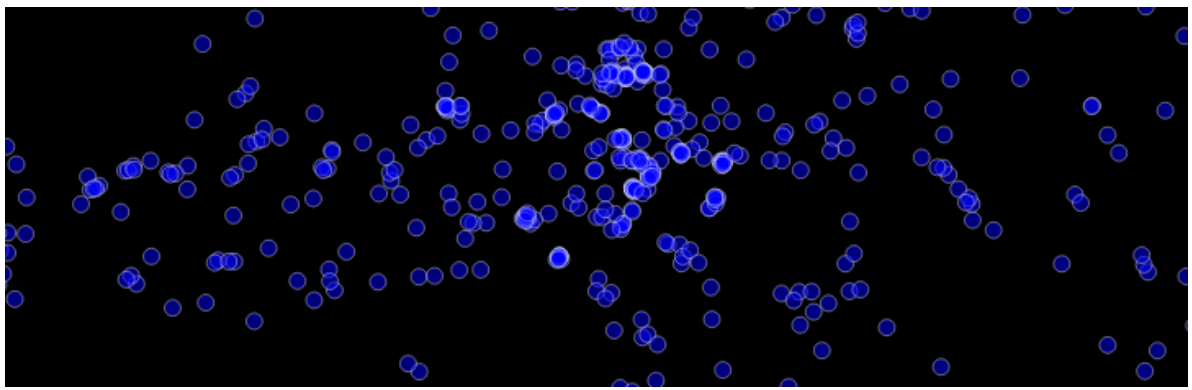
- **Scientometrics** refers to the science of analyzing and measuring a discipline's research topics, authors, publications and so forth.
- In previous works, we have proposed a framework for **spatiotemporal scientometrics** which focuses on discovering the spatiotemporal patterns of a research discipline (Hu etc., ISWC 2013; Gao etc., 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**.



# Graph and Shortest Path Distance

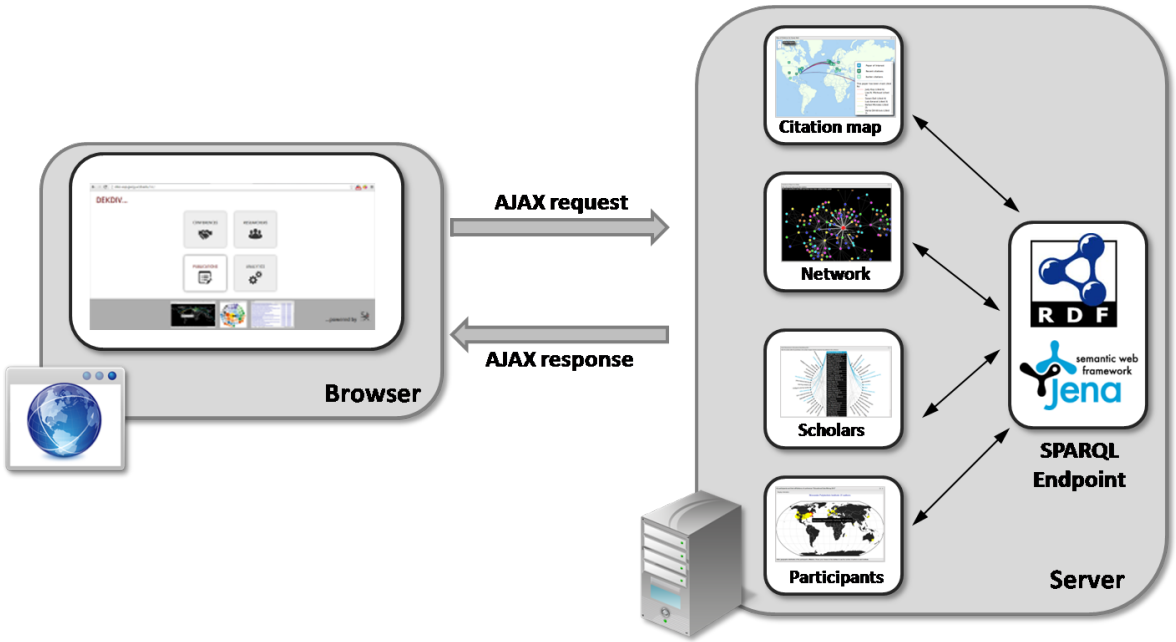
- An **academic graph** can be constructed by considering researchers as **nodes** and co-authorships as **edges**.
- From each node in the graph, we calculate its **shortest path distance** to other nodes (researchers) using Dijkstra's Algorithm.
- The shortest path distance has been used to represent how far (or how close) that two researchers are connected.
- Application: finding potential collaborators for researchers.

$$potential = sim(a_1, a_2) \times (1 - 1/d)$$

# Outline

# 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.



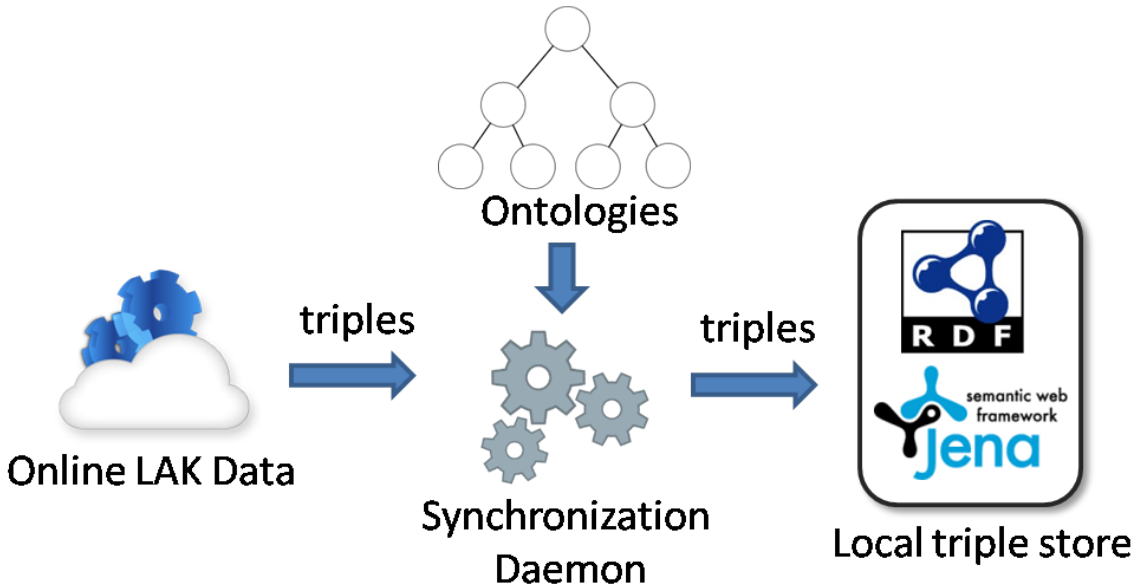
# Functional Scalability

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

```
81 <script type="text/javascript" src="js/collaborativeInstitute.js"></script>
82 <script type="text/javascript" src="js/citationMap.js"></script>
83 <script type="text/javascript" src="js/activeResearcher.js"></script>
84 <script type="text/javascript" src="js/conferenceTopic.js"></script>
85 <script type="text/javascript" src="js/paperKeywordBubble.js"></script>
86 <script type="text/javascript" src="js/mdsScholarOperator.js"></script>
87 <script type="text/javascript" src="js/coauthorTreeMap.js"></script>
88 <script type="text/javascript" src="js/paperReference.js"></script>
89 <script type="text/javascript" src="js/conferenceParticipants.js"></script>
90 <script type="text/javascript" src="js/coauthorNetwork.js"></script>
91 <script type="text/javascript" src="js/reviewerRecommendation.js"></script>
92 <script type="text/javascript" src="js/dataBrowser.js"></script>
93 <script type="text/javascript" src="js/instructionModal.js"></script>
94 <script type="text/javascript" src="js/potentialCollaborator.js"></script>
95 <script type="text/javascript" src="js/topicTrending.js"></script>
```

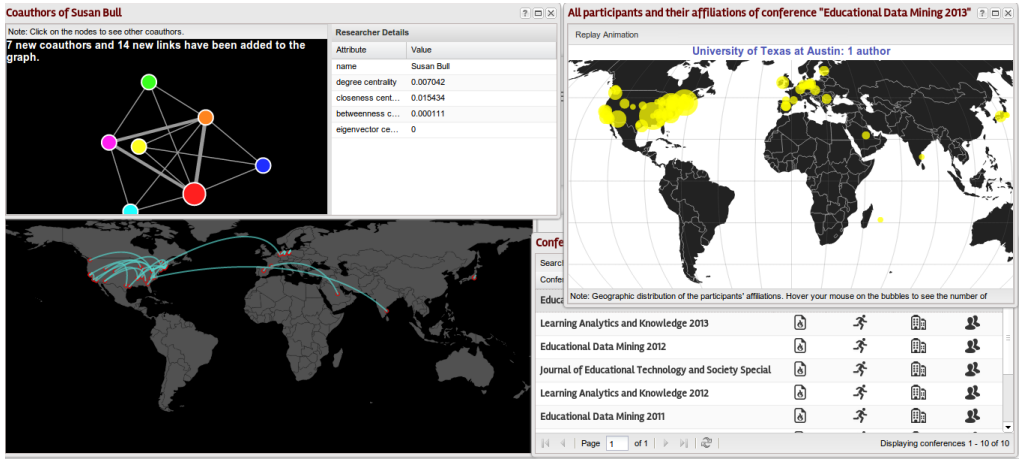
# Data Scalability

- 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.

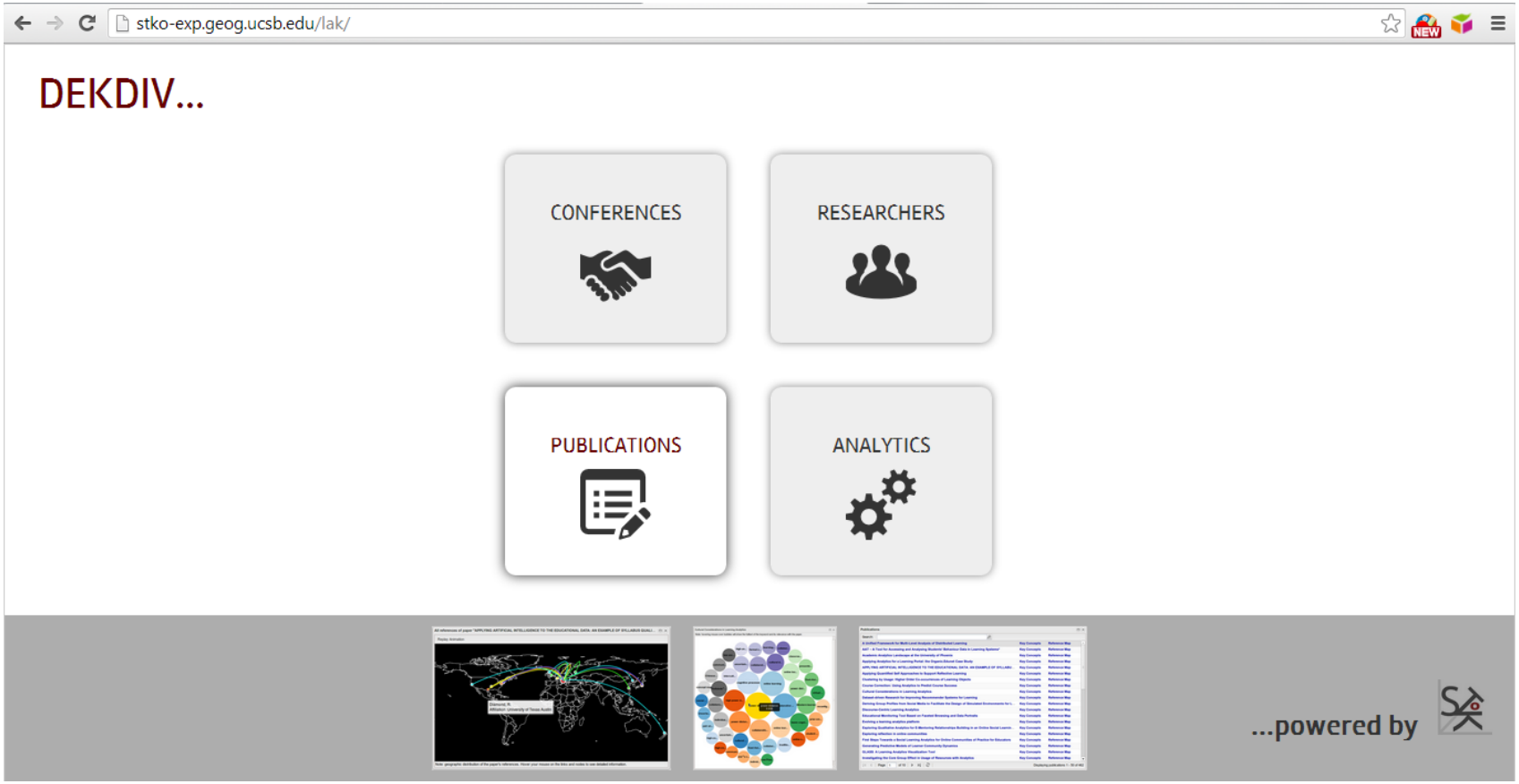


# Technology Stack

- **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

# Demo


stko-exp.geog.ucsb.edu/lak/
DEKDIV...
CONFERENCES
RESEARCHERS
PUBLICATIONS
ANALYTICS
...powered by