Analysis DBLP using Graph database

-- Data Intensive Workflow Project

Team - *Cici Huang, Kang Fang, Qihao Zhu, Xiatao Jin*
Faculty Advisor - *Jia Zhang*

Agenda

- Introduction
- Motivation
- Related work
- System design
- System implementation
- Demo
- Conclusions and future work

Introduction - Graph Database

Data Model:

- Nodes and Relationships

Examples:

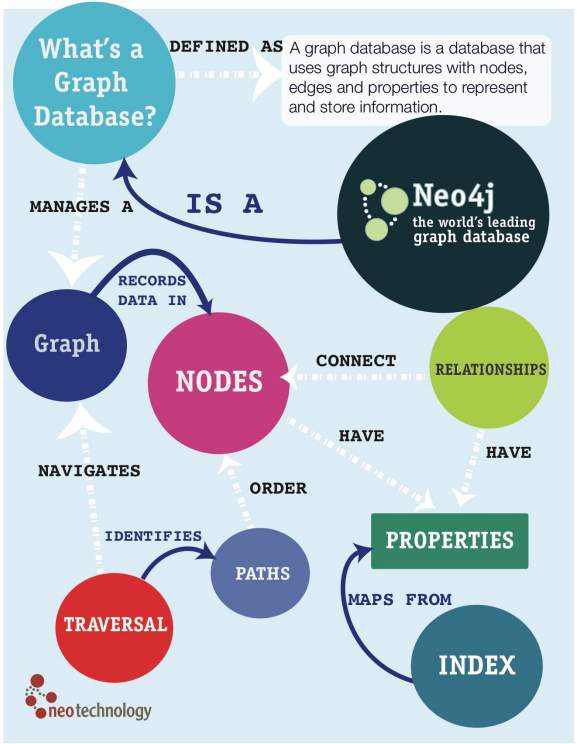
- Neo4j, OrientDB, InfiniteGraph, AllegroGraph

Pros:

- Powerful data model, as general as RDBMS
- Connected data locally indexed
- Easy to query

Cons

- Sharding
- Requires rewiring your brain



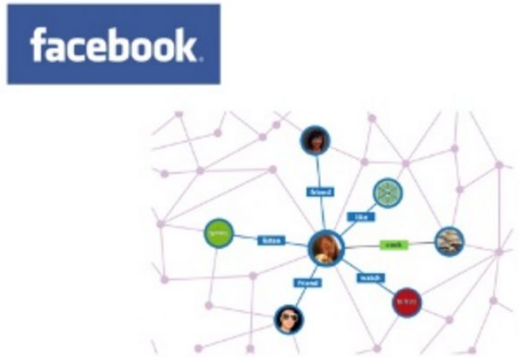
Introduction - DBLP

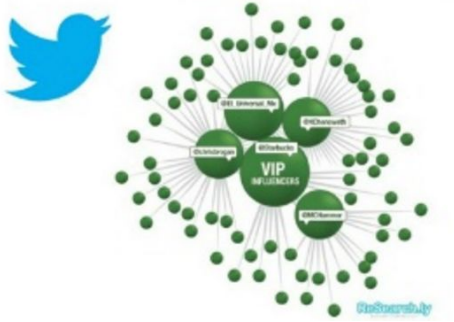
A computer science bibliography website

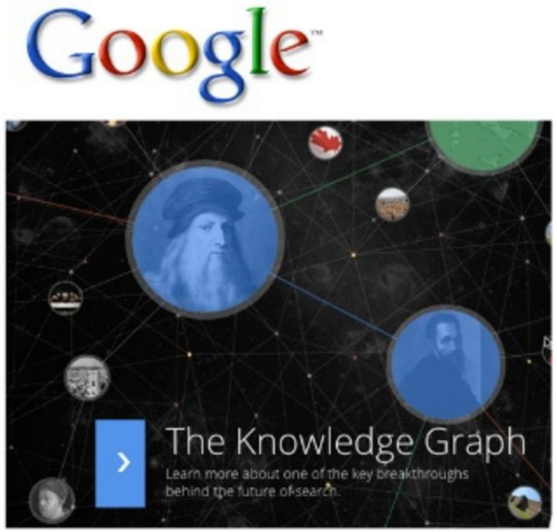
- 1.6 G +
- Sparse data
- Missing attributes

Motivation

- Graph Tech
- Neo4j is the leading graph database
- DBLP

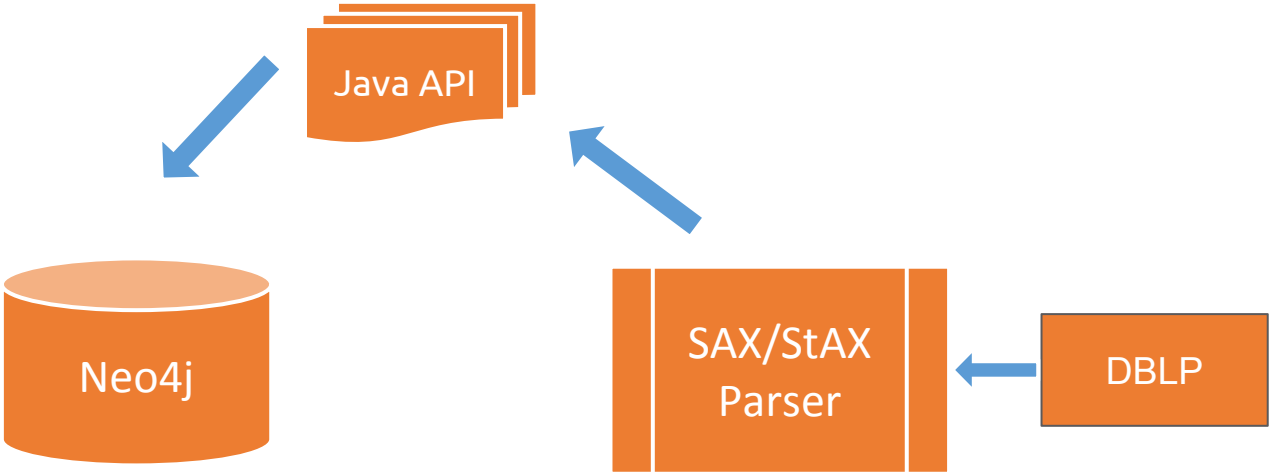




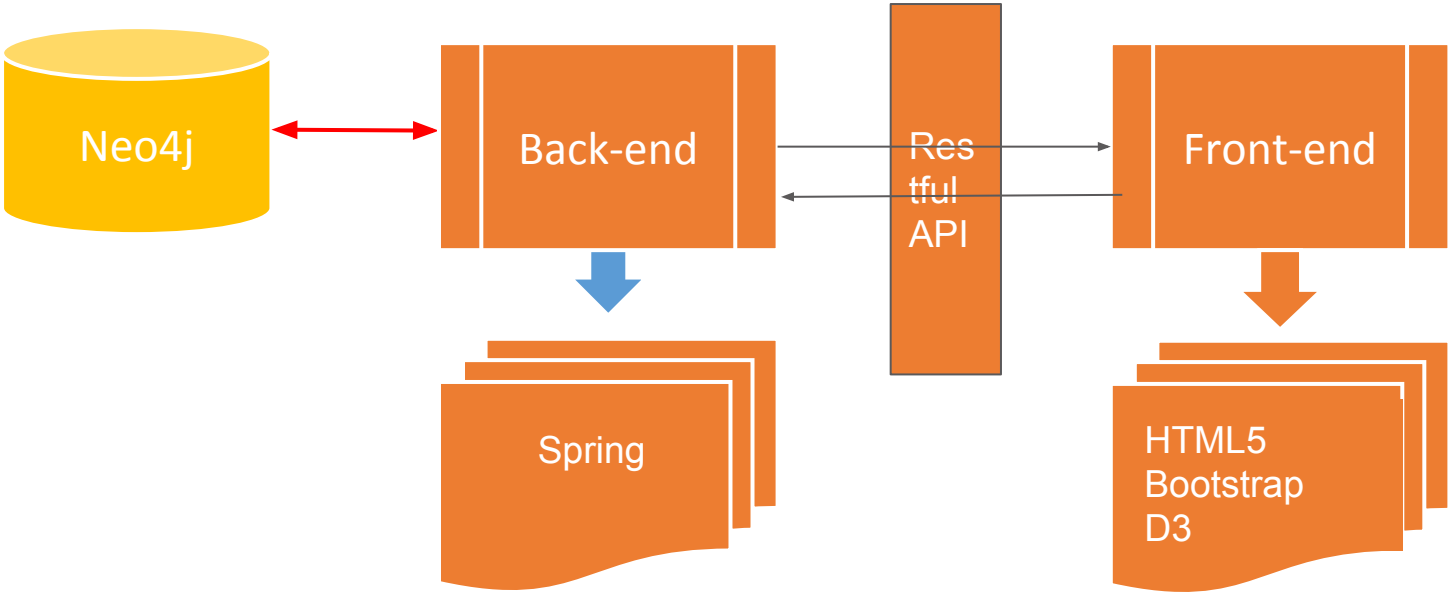


Deal with data

- 10,000+ Papers
- 7000+ Authors
- 17,000+ Relationship

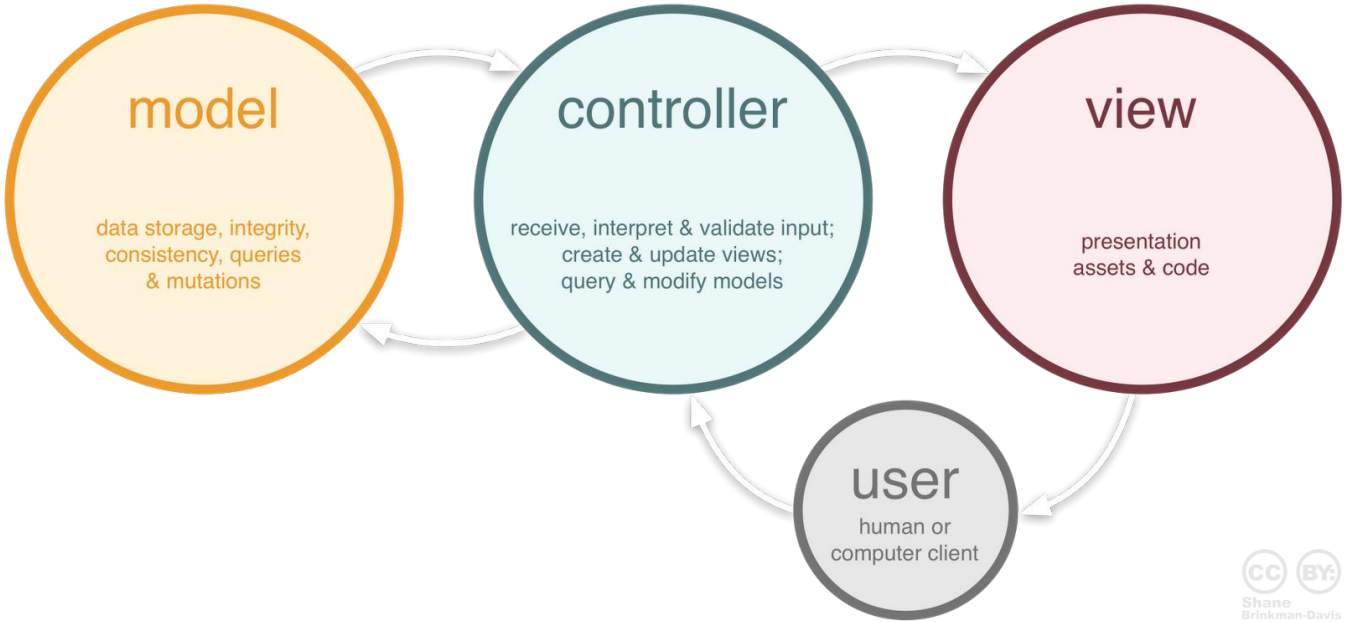


System Design



Design: Big Picture

- MVC: Model - View - Controller



Design: Big Picture

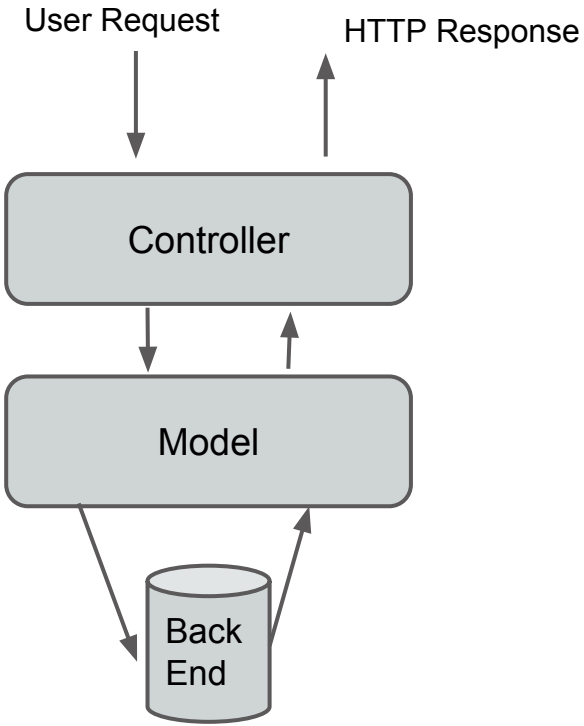
REST : Representational State Transfer

- An Architecture Style : Scalability, Simplicity, Visibility ...
- RESTful API HTTP Methods : GET, PUT, POST, DELETE
- REST in Spring Framework

Implementation: Back End

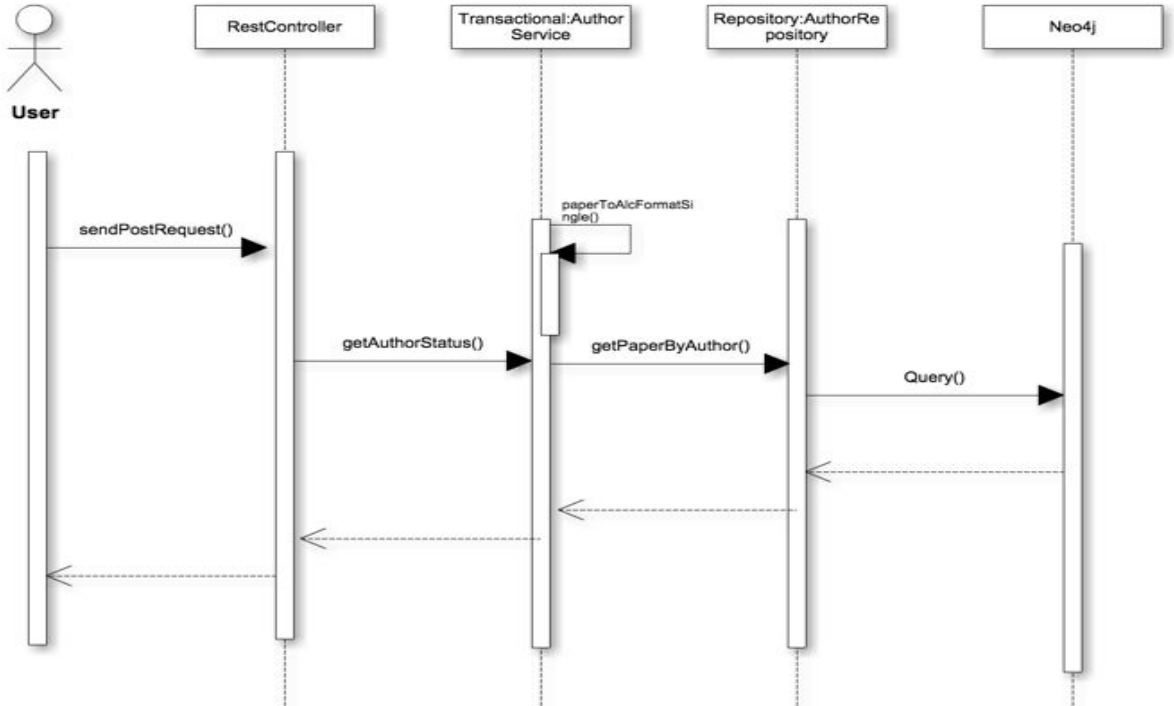
Main responsibility: providing RESTful API

Implementation: Back End



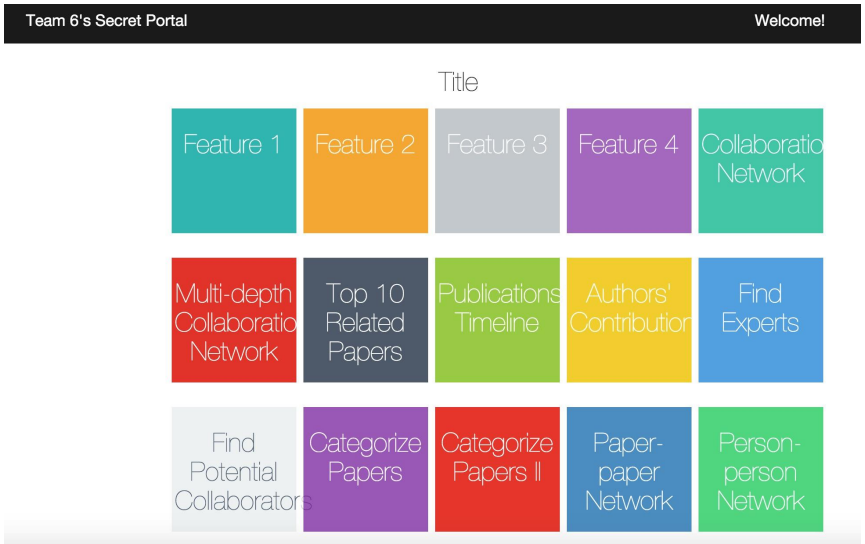
Implementation: Back End

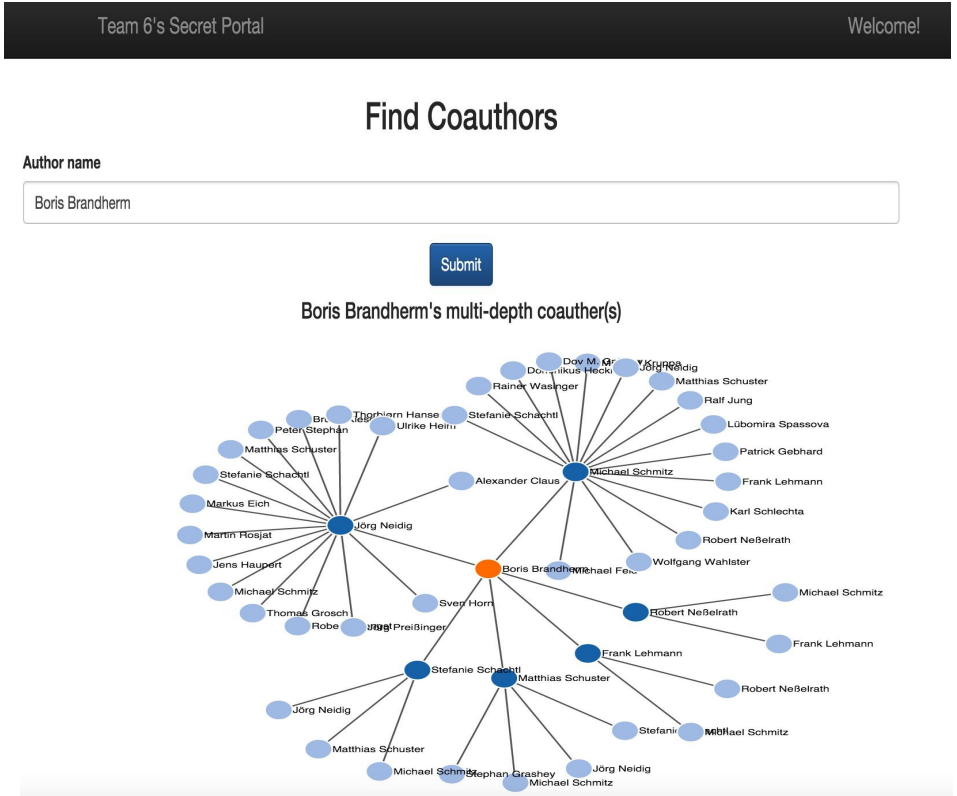
Example :/ getAuthorStatus/:authorName



Implementation: Front End

- Bootstrap
- D3





Demo

Next Step

- Performance improvement
  - Query time
  - Loading data
- Scalability
- Synchronization

Q & A