

Master Project

Various Aspects of Recommender Systems

October 24th, 2016
Master project WS 16/17

Albert-Ludwigs-Universität Freiburg

Prof. Dr. Georg Lausen
Anas Alzoghbi
Victor Anthony Arrascue Ayala



**UNI
FREIBURG**

Agenda



- Organization
- Recommender Systems
- Topics
 - Cross-domain recommendations in RecSesame
 - Scientific Paper recommendation
- Agenda for next week

Requirements



- Study regulations (Studienordnung)
 - 16 ECTS → 480 hours
- Master project
 - Team size: 1-3 students
 - Project report: 40 pages
 - Short presentations: 2-3 (individual as needed)
 - Final presentation: 25 min
- Some preconditions
 - Prior knowledge in Java programming
 - Recommended lecture “Data Analysis and Query Language”

Organization



- Time & Place
 - Monday 14-16 (c.t.)
 - Geb. 51 – SR 01 029
- Website (<http://dbis.informatik.uni-freiburg.de>)
 - Apply via [HISinOne](#)
 - Startseite › Lehre › Lehrangebot › Wintersemester2016/17 › Various Aspects of Recommender Systems

General goals



- Collective work on a project
- Gain experience in research and development method
- Improve individual programming skills
- Incorporate in new topics (Semantic Web, Recommender systems,...)
- Learn about problems of larger projects



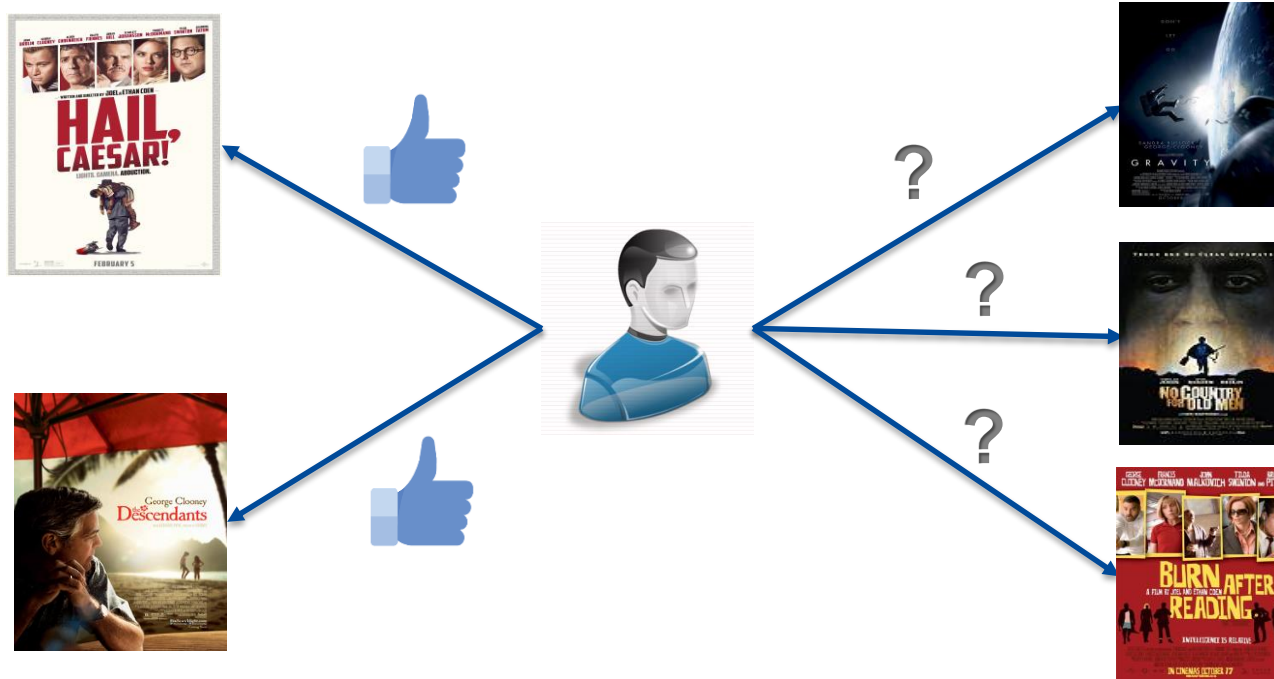
- Workload of every student must be clearly distinguishable
- Some Criteria
 - The scope and difficulty of the work / implementation
 - Individual contribution
 - Team performance: a successful project has a positive effect
 - Role and participation in the team (coordination, etc.)
 - Quality of code (formatting, documentation)
 - Individual report (project report)
 - Presentations (especially the final presentation)

Master projects

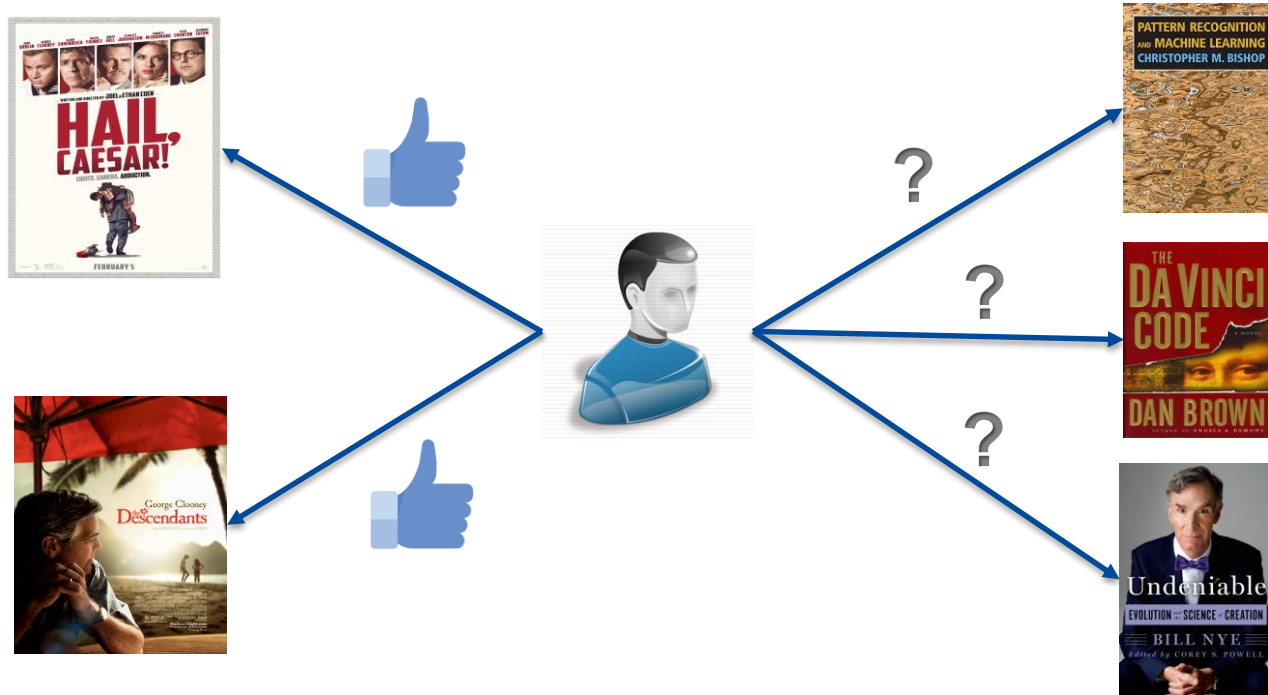


1. Cross-domain recommendations on RecSesame (Anthony)
2. Mining and integrating conference meta-data (Anas)

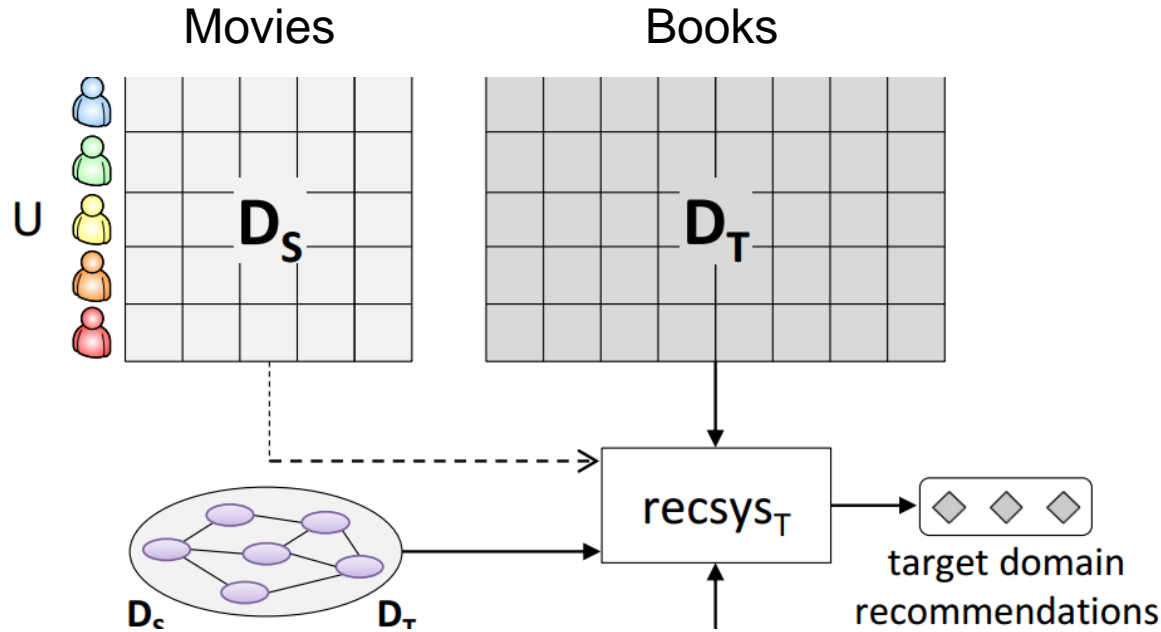
Cross-domain RS- 1st project



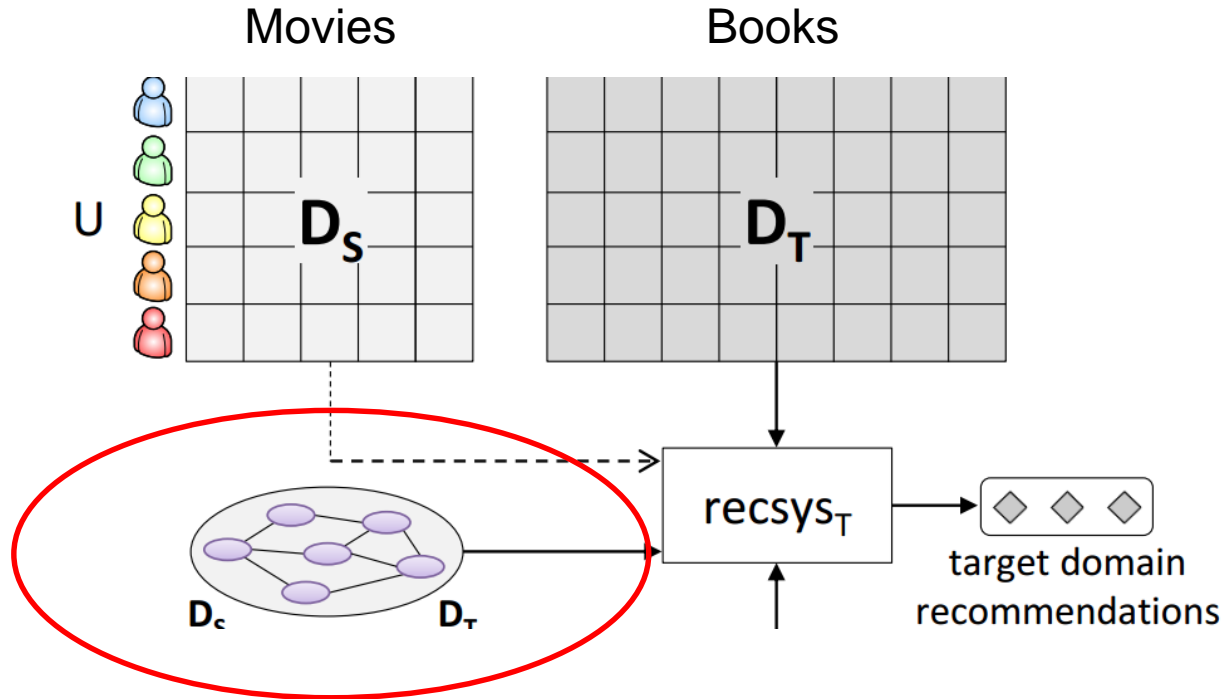
Cross-domain RS



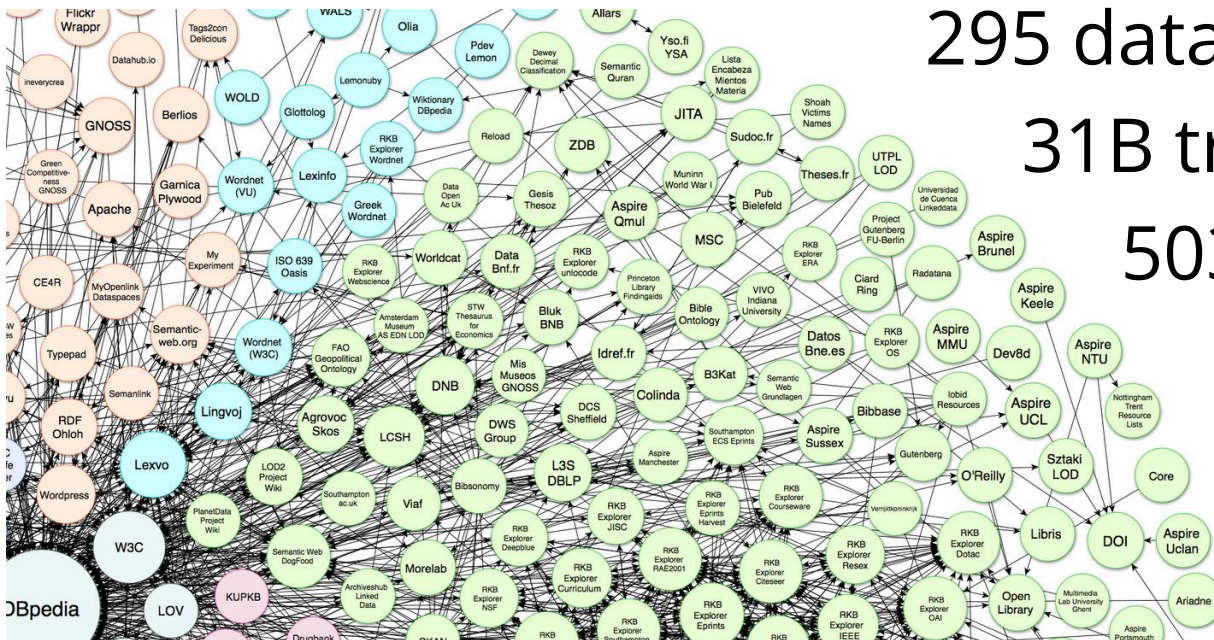
Knowledge Linkage



Knowledge Linkage



Linked open Data



295 datasets
31B triples
503M out links

Partial view of Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak.

Linked open Data



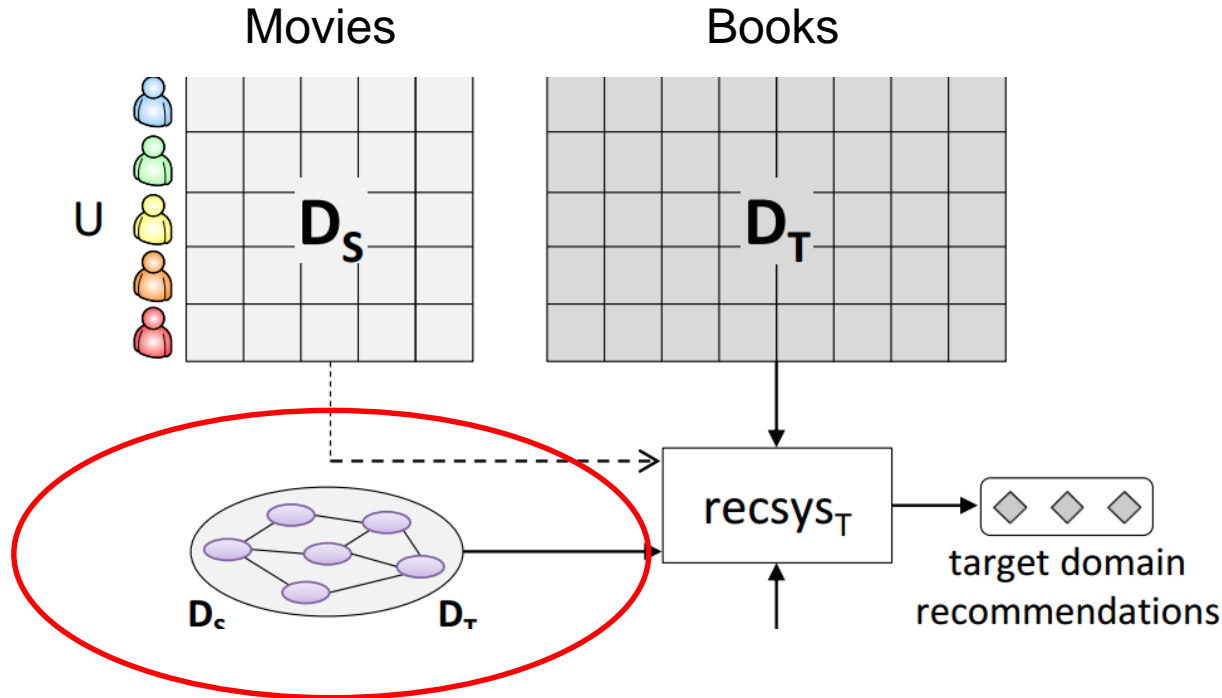
- Example:

- <http://www.visualdataweb.org/refinder/refinder.php>

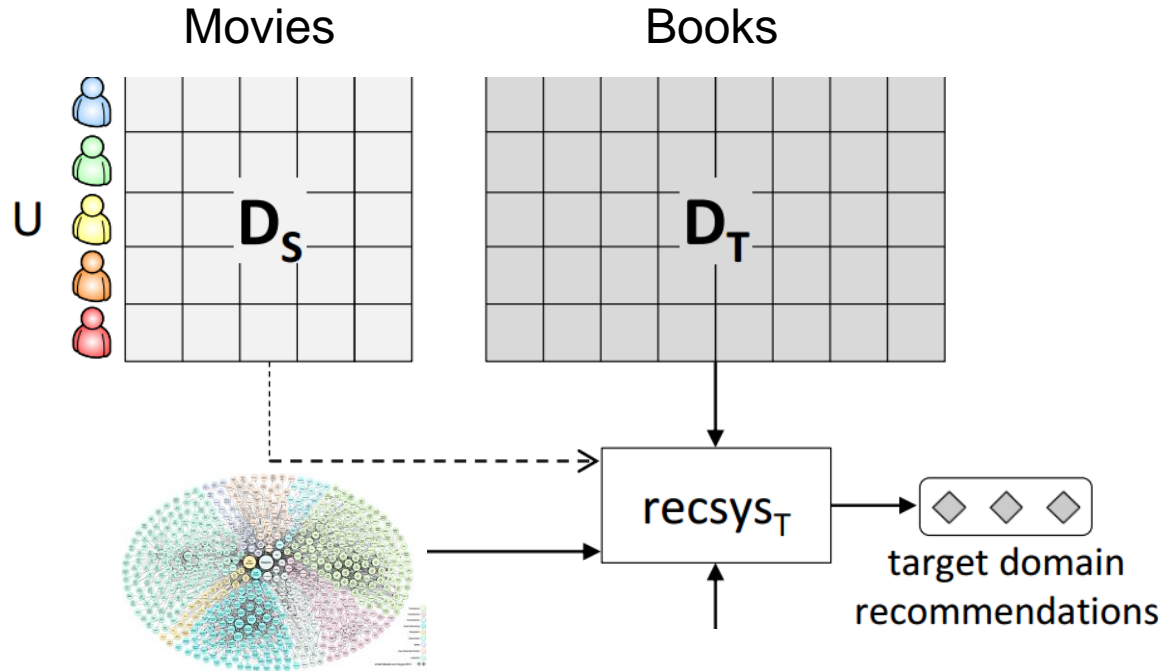


Partial view of Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak.

Knowledge Linkage



Knowledge Linkage



The project



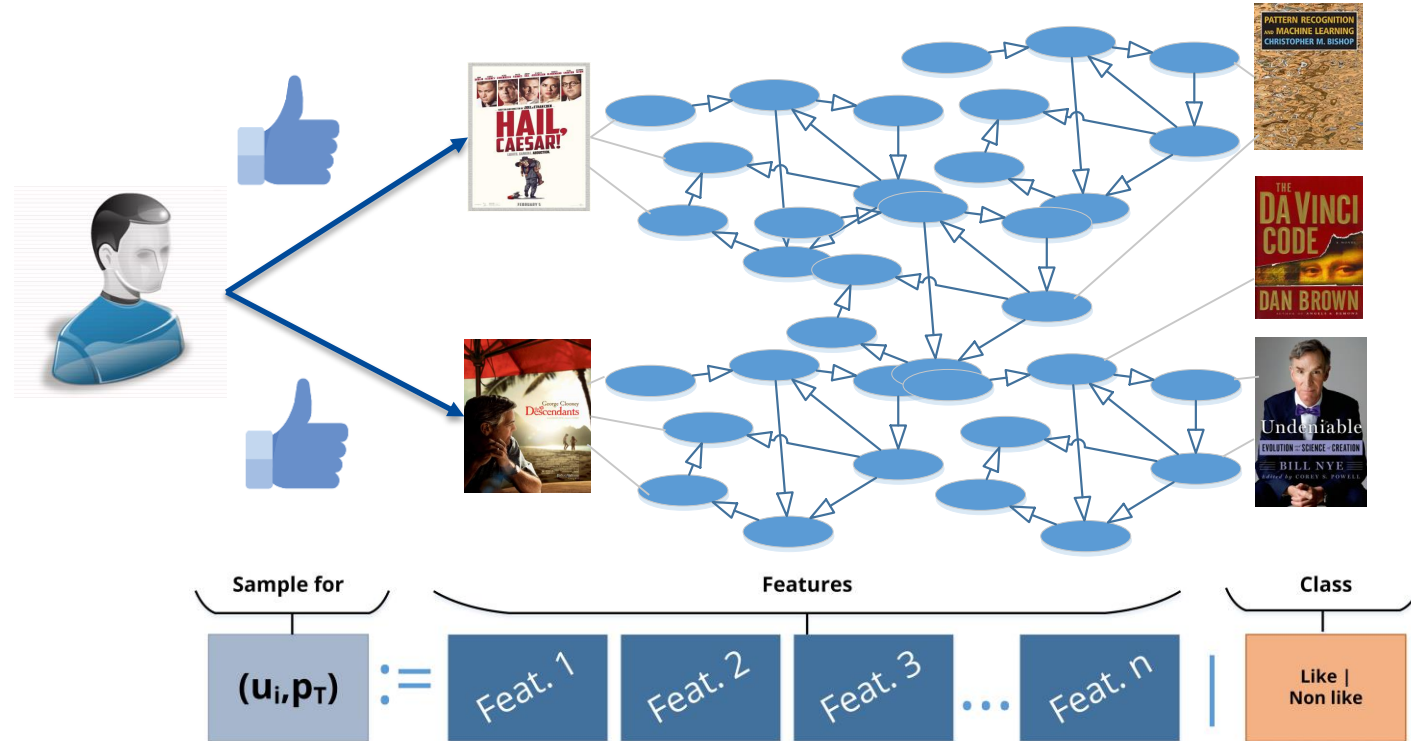
1. Understand the data
2. Design a cross-domain RS
3. Integrate the recommender into RecSesame
4. Evaluate the recommender

Data

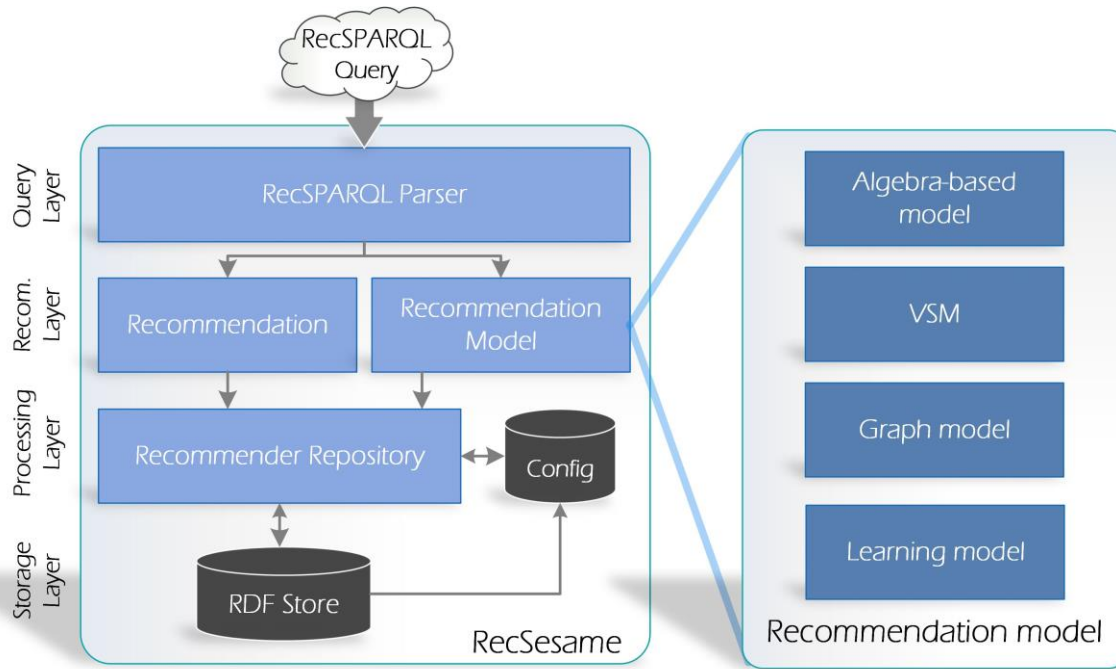


- Collection of Likes (Facebook)
 - Domains: music, movies, books
- Challenge ESWC'15
- We extracted data from Dbpedia
- Items are interconnected

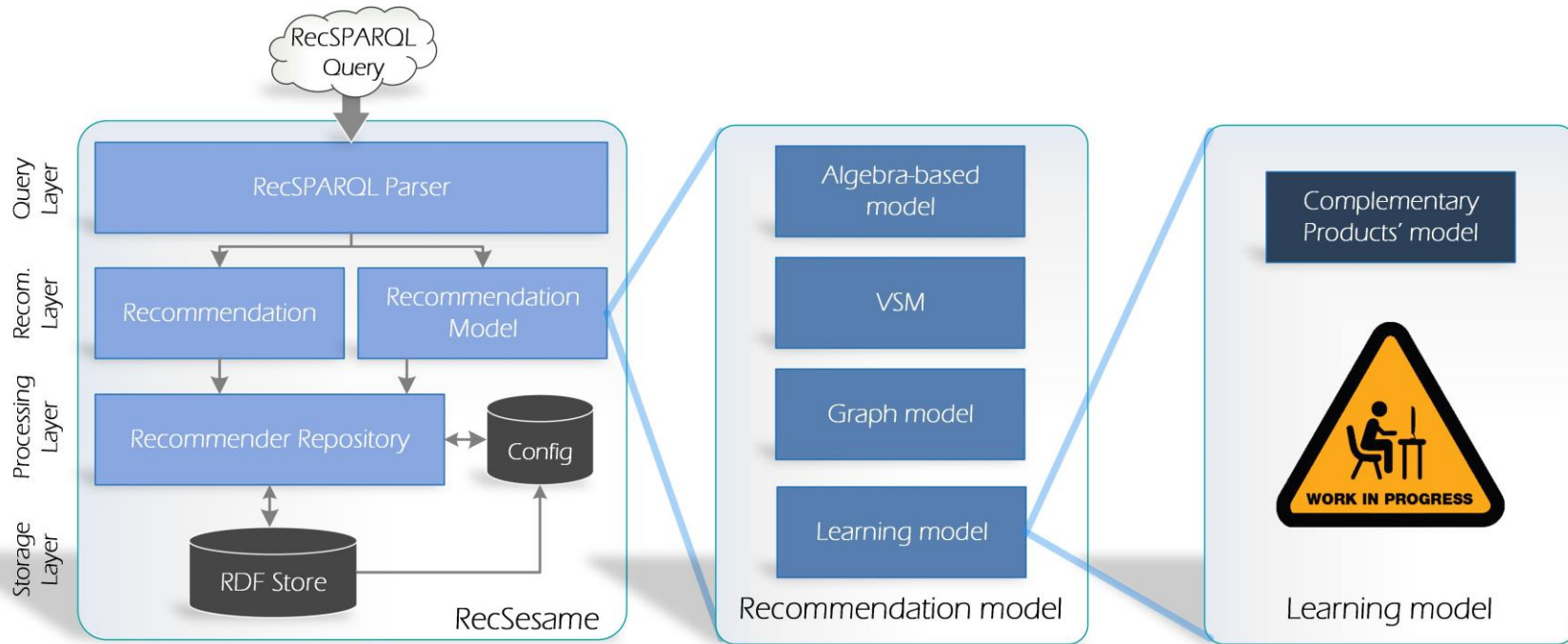
Design



RecSesame - models



RecSesame – learning models



Evaluation – supported metrics



- Ranking Metrics
 - Precision, Recall, F-Measure, Mean Reciprocal Rank, Normalized Discounted Cumulative Gain, Area Under Curve
- Prediction Metrics
 - Mean Absolute Error, Root Mean Squared Error

Submission of task (compulsory)



- 2 teams, 2 students each
- Deadline: 07.11.2016
- Pre-requisite to participation

Submission of task (compulsory)



1. Get started with RecSesame
 - Submit evaluation results for small dataset
2. Implement a dummy recommender and evaluate it
3. Report
 - Design proposal (1 page)
 - Related work (3 pages)

Scientific Paper recommendation- 2nd project



- Recommend Scientific papers to scholars
- Content-Based recommendation
- Publication history
- Exploiting publicly available meta-data
 - Title
 - Abstract
 - Keyword list
 - Publication year

Scientific Paper recommendation



- For a researcher (r)
 - m Publications
 - n Keywords
 - p' : A candidate paper

		k_1	k_2	k_3	...	k_n
Previous/relevant publications	p_1	$\rightarrow w_{1,1}$	$w_{1,2}$	$w_{1,3}$...	$w_{1,n}$
		
	p_m	$\rightarrow w_{m,1}$	$w_{m,2}$	$w_{m,3}$...	$w_{m,n}$
Candidate papers	p'	$\rightarrow w'_1$	w'_2	w'_3	...	w'_n
	p''	$\rightarrow w''_1$	w''_2	w''_3	...	w''_n
		

Mining and integrating conference meta-data



- Extracting structured information of conference papers
 - Title
 - Abstract
 - Authors
 - Keywords list
 - Year
 - Date & time
 - Pdf file
- Design a tool for an existing recommendation app that can generically deal with various conferences websites
- Mine meta-data for
 - Main track papers
 - Workshops papers
- Integrate existing tools/systems
 - [jsoup](#)
 - [GROBID](#) (for pdf files)

Mining and integrating conference meta-data



- Example of Conference websites:
 - <http://iswc2016.semanticweb.org/pages/program.html>
 - <http://2016.eswc-conferences.org/program>
 - http://iswc2014.semanticweb.org/program_glance.html
 - <http://iswc2015.semanticweb.org/program>
 - <http://www.www2015.it/program/>
 - <http://www2016.ca/program-at-a-glance.html>
 - <http://2014.eswc-conferences.org/program/accepted-papers.html> (pdf files)
 - <http://2015.eswc-conferences.org/program/accepted-papers>

Recommendation web app - demo



Jsoup

<https://try.jsoup.org/>

ISWC 2015

URL: <http://iswc2015.semanticweb.org/papers>

CSS selectors:

List tag: .views-field
Paper Name: strong
Author: em
PDF: a[href]

ESWC 2014

URL: <http://2014.eswc-conferences.org/program/accepted-papers.html>

CSS selectors:

List tag: .field-item li
Paper Name: em
Authors: span

ESWC 2015

URL: <http://2015.eswc-conferences.org/program/accepted-papers>

CSS selectors: same as ESWC 2014

Submission of task (compulsory)



- Team of 2 students (1 student is also accepted)
- Deadline: 07.11.2016
- Pre-requisite to participation

Submission of task (compulsory)



1. Get started with [Jsoup](#), [GROBID](#)
2. Implement a dummy crawler for [ISWC 2015](#)
 1. Crawl paper names, authors, pdf files, and abstracts from pdfs
3. Report
 - Design proposal (1 page)
 - Related work (3 pages)
report on existing systems/tools/methods to solve this problem

Thank you!



**UNI
FREIBURG**

Any questions?