



MOOCdb: Developing Data Standards for MOOCs

Srećko Joksimović

s.joksimovic@ed.ac.uk

@s_joksimovic

Kalyan Veeramachaneni

kalyan@csail.mit.edu

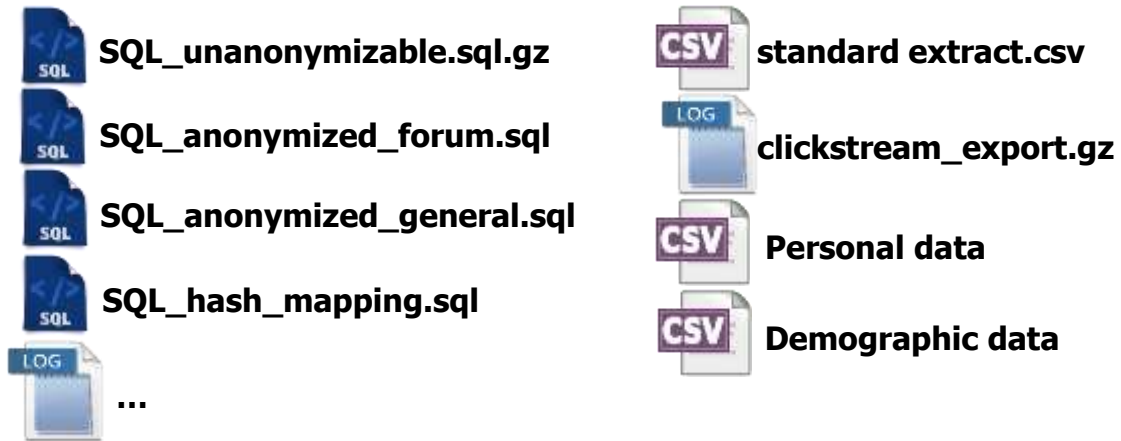
Dragan Gašević

dragan.gasevic@ed.ac.uk

FutureLearn Academic Network Conference

15 June 2015













Data sources



Data sources



Weekly data packages

-  **auth_user-`{site}`-analytics.sql**
-  **auth_userprofile-`{site}`-analytics.sql**
-  **certificates_generatedcertificate-`{site}`-analytics.sql**
-  **course_structure-`{site}`-analytics.json**
-  **courseware_studentmodule-`{site}`-analytics.sql**
-  **email_opt_in-`{site}`-analytics.csv**
-  **student_courseenrollment-`{site}`-analytics.sql**
-  **user_api_usercoursetag-`{site}`-analytics.sql**
-  **user_id_map-`{site}`-analytics.sql**
-  **`{org}`-`{course}`-`{date}`-`{site}`.mongo**
-  **wiki_article-`{site}`-analytics.sql**
-  **wiki_articlerevision-`{site}`-analytics.sql**

Daily data packages

-  **`{org}`-`{site}`-events-`{date}`.log.gz.gpg**

Challenges

- Analytics across several courses
- Analytics across different platforms
- Analytics across different institutions
- Sharing data

Solution?

- Collaborative data science platform
 - Standardize data storage
 - Generalizable across courses and data providers (currently OpenEdX, edX and Coursera)
 - “Data being shared without data being exchanged”
 - Sharing and reproducing the results

MOOC data science commons

MOOCDB



Translation
scripts

Tools

Raw data sources

coursera

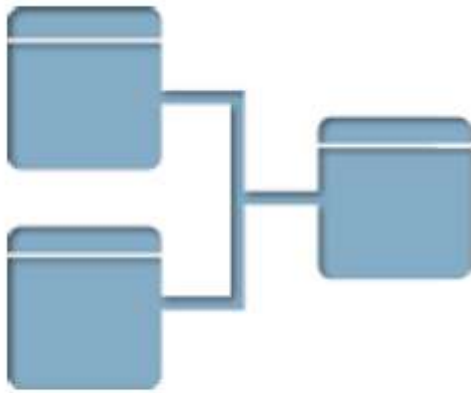
Translate



Translate

MOOCdb

<http://moocdb.csail.mit.edu/wiki/index.php?title=MOOCdb>



Shared data model

User information

- User PII table
- Global User table
- Course User table

Observing mode

- Observed Events table
- Resources table
- Resources Types table
- URLs table
- Resource URLs table

Submitting mode

- Problem Types table
- Problems table
- Submissions table
- Assessments table

Collaborating mode

- Collaborations table
- Collaboration Types table

Feedback mode

- Feedbacks table
- Questions table
- Answers table
- Surveys table

Collaborative platform and applications



Translate

edX

coursera

Translate

Coursera



MOOCdb doc



Github repo

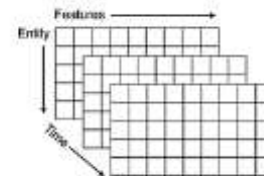


MY FEATURE FACTORY

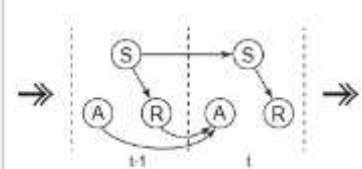
Feature factory

LabelMe

LabelMe



Digital learner quantified



Problem analytics



MY MOOCViz

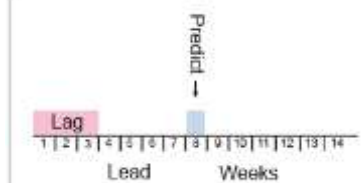
My MOOCViz



Social network analysis



Forum analysis

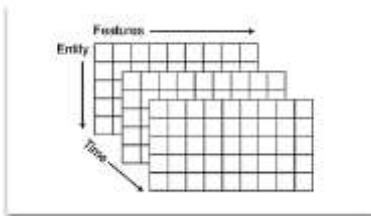


Dropout prediction

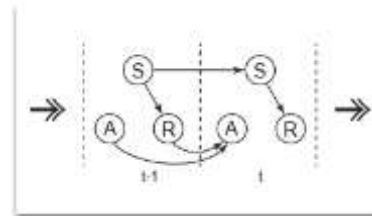
Current state

- Established network of institutions
 - MIT, Stanford, University of Michigan, University of Edinburgh, University of Queensland, University of Texas (Austin)
- Release of open source software
- Development and release of the first data analytics framework

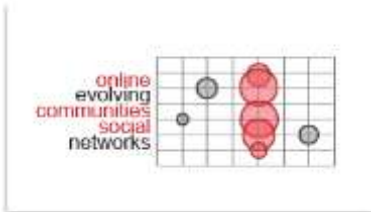
Next steps



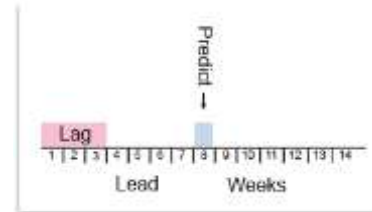
Digital Learner Quantified



Problem analytics



Discussion forum analysis



Dropout prediction

LabelMe

LabelMe



Social network analysis

Collaboration

- If you are interested in...
 - Development
 - Feature modeling
 - Translating your data
 - Testing

kalyan@csail.mit.edu

s.Joksimovic@ed.ac.uk



MOOCdb: Developing Data Standards for MOOCs

Q&A

Srećko Joksimović
s.joksimovic@ed.ac.uk
@s_joksimovic

Kalyan Veeramachaneni
kalyan@csail.mit.edu

Dragan Gašević
dragan.gasevic@ed.ac.uk

FutureLearn Academic Network Conference
15 June 2015