

Analyzing OpenCourseWare usage by means of social tagging

Julià Minguillón^{1,2}, Jordi Conesa²

¹UOC UNESCO Chair in e-Learning

²Universitat Oberta de Catalunya

Barcelona, Spain

Table of contents

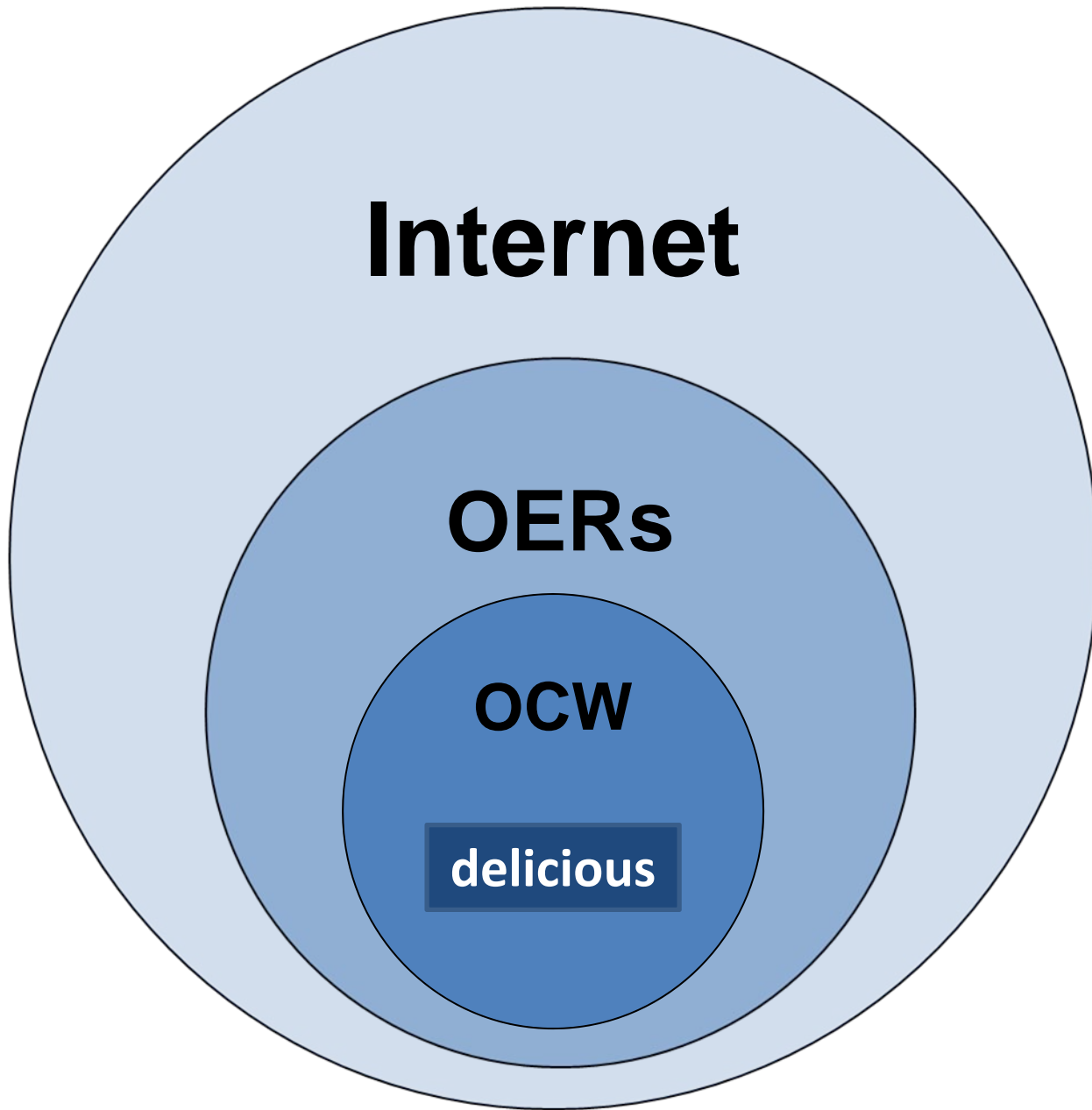
- Open Educational Resources
- Social tagging
- Exploratory analysis
- Preliminary results
- Discussion
- Conclusions
- Future research

Open Educational Resources

- Everybody is creating and publishing OERs
- Available through repositories, websites, ...
 - MERLOT, ...
 - OpenCourseWare
- But...
 - Poor metadata descriptions, flat structure
 - Wide degree of granularity
 - Who is (re)using them?

Searching at OCW

- Limited possibilities:
 - Filter by keyword
 - Language
 - Source (OCW site)
- No shared taxonomies between OCW sites
- **Goal:** improve users' support when browsing and searching for OERs



Social tagging

- Web 2.0 service (a.k.a. social bookmarking)
- Organize, describe and share URLs
- Triplets: URL, user, {tags}
- Delicious:
 - Largest service of its kind
 - Sustainability: Yahoo wants to close/sell delicious
- Simple poll: how many of you use delicious?

opencourseware

Search

- mit opencourseware
- harvard opencourseware
- mit opencourseware course materials
- ocw opencourseware
- opencourseware offers online
- opencourseware consortium forum

Filter by Tag:

- education
- learning
- free
- opencourseware
- Double-click to add a tag

Bookmarks Saved From: oldest bookmark to now - 16848 Results

Times are GMT

1D | 1M | 6M | 1Y | Max |



Show: Everybody's bookmarks (1000+)

Sign in to search your own bookmarks

Everybody's bookmarks

16,848 results

Sponsored Results

Open Courseware
 Move your career forward with an accredited online degree!
www.CourseAdvisor.com

OpenCourseWare Consortium

ocwconsortium.org/

4587

education learning opensource free opencourseware

Free Online Course Materials | MIT OpenCourseWare

ocw.mit.edu/DownWeb/web/home/home/index.htm

14888

education mit learning free online

Academic Earth - Video lectures from the world's top scholars

academicearth.org/

25238

education video lectures learning university

OpenCourseWare Finder

openccontent.org/ocwfinder/

1446

education learning free reference opencourseware

Everyone's Bookmarks for:
OpenCourseWare Consortium
 oocwconsortium.org/

History Notes

Saved 4587 times, first saved by Djembe on 22 Jun 06. View Chart

- 10 APR 11 [brettsev](#) education opensource reference pedagogy
- 08 APR 11 [everthonVS](#) elearning open courses courseware

[Elisabeth Wulff sahlén](#) opencourseware free elearning
- 06 APR 11 [mk](#) education free elearning resources oers

OCW Consortium, eta bere plataforman mundo osoko hezkuntza-edukiak eskuragarriak eta berrerabiltzeko prest eskaintzen du

[laura da rocha](#) sarbide_askea

A free and open digital publication of high quality university-level educational materials. Open Access unibertsitate-mailako irakaskuntzarako publikazio digitalak.

[Goizane S. Aviles](#) en askea musutruk digital publikazio materiala irakaskuntza unibertsitatea openaccess

Ikaskuntzarako biltegi digitala: OCW

[Xixili Abal](#) openaccess ikaskuntza

mundo osoko hezkuntza-edukiak eskuragarriak eta berrerabiltzeko prest eskaintzen du.

[Danel Miner](#) openaccess

Nazioarteko erakunde patzuergo bat dago, OCW Consortium, eta bere plataforman mundo osoko hezkuntza-edukiak eskuragarriak eta berrerabiltzeko prest eskaintzen du.

[enara aldeiturriaga...](#) education learning free e-learning online

- [Save this bookmark](#)
- [Share this bookmark](#)
- [Look up another URL](#)

Tags

▼ Top Tags

education	2338
opencourseware	1538
learning	1294
opensource	1065
free	1041
elearning	952
e-learning	783
online	683
courses	477
resources	449
reference	371
ocw	330
university	295
courseware	190
academic	187
open	186
school	179
course	157
college	156
books	142
teaching	140
science	137
lectures	123
community	120
courses	96
oer	93
mit	92

What / How do users tag?

- Top level web sites
- Specific links (i.e. courses)

- Web site: mit, ocw
- Content: math, course, pdf
- Usage: phd, ds106, toread

- If delicious is used for content organization, then tags become summarized descriptions

Basic idea

- What if we obtain “all” tags from “all” users sharing “all” the URLs for a given concept?
- Tags could reveal a hidden semantic structure used for organizing and describing content
- What does “OCW” mean for OCW users?

Methodology

- Retrieve “all” triplets for a given search term (opencourseware + ocw)
- Do some data cleansing
- Select the most important triplets
- Build a binary data set (N x T):
 - URL X tagged by user Y with tags T_1, \dots, T_T
- Perform some exploratory data analysis
- Interpret results

Retrieving data from delicious

- scripting + wget + HTML parsing
- Problem: IP throttling
- Slow, “risky”, tricky process
- Search term: “opencourseware+OR+ocw”
- Before data cleansing and selection:
 - 2479 URLs
 - 50 pages of users’ tags per URL
 - 3 GB of information

Data cleansing

- Plurals, capital letters
- Misspelling: course, coruse
- Multilingualism: curso, curs, kurs, ...
- Equivalent tags: math = mathematics
- Abbreviations: cs = compsci = computer science
- Compound terms: web + 2.0 = web20
- Spammers
- URLs with no tags

Triplet selection

- Remove those users tagging less than M resources (spammers, occasional users)
- Remove those URLs tagged by less than N users (spammers, broken URLs)
- Keep only a fraction of tags (P) which are the most used (Pareto-like)
- $M=5-10$, $N=5-10$, $P=80\%-90\%$

Resulting data sets

- 389866 triplets (N):
- Before: 1456 URLs, 26753 users, 21384 tags
- After: **1297 URLs, 24981 users, 434 tags (T)**
- Course subset: course URLs at MIT OCW
 - 17024 triplets
 - **351 URLs, 5689 users, 362 tags (120 if P=80%)**
- Data set – subset (Non-course):
 - **946 URLs, 12290 users, 412 tags (116 if P=80%)**

Most used tags

Non-course data set

- **education**
- **resource**
 - **free**
 - **learning**
- elearning
 - **video**
- teaching
 - **science**
- lessonplan
 - web20
- **reference**
- technology
 - online
 - **course**
 - **lecture**

Course data set

- mit
- **education**
- opencourseware / ocw
- **course**
- programming
- **video**
- **learning**
- **lecture**
- math
- **science**
- python
- algorithm
- **free**
- **reference**
- **resource**

Exploratory data analysis

- **Goal:** analyze how tags are used
- R 2.12.1 + psych package
- Factor analysis:
 - F factors (eigenvalue ≥ 1 , $F \ll T$)
 - Maximum likelihood
 - Varimax rotation
 - Only factors with weight ≥ 0.1 are considered (**0.3**)
- Problems: instability, too many parameters

Preliminary results

- 20.1% of variance explained (20 factors, 16.7%)
- Interesting factors (out of 50):
 - ML11: **programming, algorithm, tutorial**, howto
 - ML17: **algebra, linear**, math, linearalgebra
 - ML21: **scheme, lisp, sicp**, programming, book
 - ML29: **psychology, neuroscience, brain, cognitive**
- Using P=90% (361 tags) also generates:
 - ML31: politics, society
 - ML39: design, usability, ui, hci, interface

Discussion

- Tags related to web sites are too general...
- ...but both data sets share $\approx 50\%$ of tags
- General (i.e. most used) tags are not relevant for factoring purposes (mit, ocw, course, ...)
- Factors group tags semantically related, mostly describing the domain knowledge
- $P=80\%$ seems to produce more stable results, but interesting factors are missed (i.e. HCI)

Conclusions

- OCW could take advantage of “advanced” users tagging its courses
- Non “tech” domains are underrepresented
- Data cleansing is critical
- Proposal: describe domain (keywords), type of resource (book, course) and format (pdf, video)
- Some parameters need fine-tuning
- More experimentation is needed!

Future research

- Automate and improve data cleansing
 - Use of Wordnet / DBpedia / ontologies
- Non-Negative Matrix Factorization + clustering
- Build recommendation systems:
 - Tags / Related resources
- Build reputation schemes:
 - Most expert users
- Pilot at <http://ocw.uoc.edu>

Thank you!

Julià Minguillón

jminguillona@uoc.edu

