



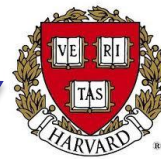
Linking Data Across Universities: An Integrated Video Lectures Dataset

Miriam Fernandez, Mathieu d'Aquin, Enrico Motta
October, ISWC 2011

- Need to integrate and share educational material across institutions

Any material about global warming?...

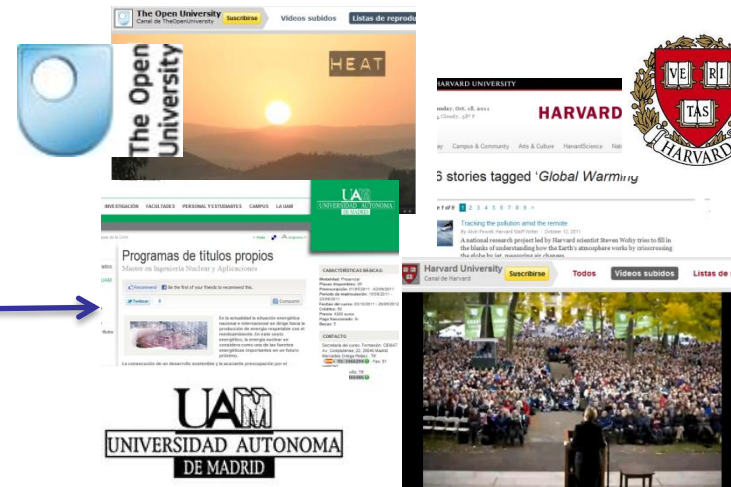
Uff... Where to start?!



- Interlink educational information across universities through the use of LD principles and technologies
- Create a consolidated dataset of video material



One unique
information
space



- Use Linked Data
 1. Select and extract educational information from various sources
 2. Reuse well-known vocabularies to describe and structure the previously extracted data
 3. Integrate the educational information in a common categorization scheme

- Videlectures.net

Title → Collective Semantics: Collective Intelligence & the Semantic Web - Semantically enriching folksonomies with flor

Author → author: Sofia Angeletou, The Knowledge Media Institute, The Open University
 published: Aug 14, 2008, recorded: June 2008, views: 509

Published and recorded dates →

Categories → Top » Computer Science » Semantic Web

Associated slides ↓

Video content → 

Description → Web 2.0 has introduced new style of information sharing featuring mass user participation, social networking, heterogeneity of data sources, and a huge scale of information and knowledge, posing difficulties in discovering relevant information. The Semantic Web may contribute by providing a language basis and ontologies to support structuring, or introducing new ways to explore the information space. This may be achieved by combining semantics from semantic web resources with structure of the

ESWC'08 - Tenerife Workshops

Semantically Enriching Folksonomies with

Sofia Angeletou, Marta Sabou and Enrico Molta

NeOn KMi The Open University

Slides

- 0:00 Semantically Enriching Folksonomies with
- 0:08 Semantic Web2.0
- 0:24 Web2.0
- 0:41 tagging systems' characteristics
- 1:24 ..an example
- 1:36 .. some missed photos
- 1:57 modifying the query..
- 2:07 our goal
- 2:27 our goal
- 3:05 our goal
- 3:37 "Dolphin OR Seal OR Sea Elephant OR Whale"
- 3:43 existing work on folksonomy enrichment
- 4:31 our approach.

- 25 YouTube university channels

```
<entry gd:etag='W/"DkADSH47eCp7ImA9WhZWFEg."'>
  <id>tag:youtube.com,2008:video:zZCaHSW88Ts</id>
  <published>2011-02-18T11:41:08.000Z</published>
  <updated>2011-05-15T10:19:39.000Z</updated>
  <category scheme='http://gdata.youtube.com/schemas/2007/categories.cat'
    term='Education' label='Education'/>
  <category scheme='http://gdata.youtube.com/schemas/2007/keywords.cat'
    term='Dr Barry Cooper'/>
  <title>Intro to Professional Practice (Children & Families)</title>
  <author> ... </author>
  <media:description ... </media:description>
  <media:keywords>...</media:keywords>
  <media:thumbnail .../>
  <yt:duration seconds='399'/>
  <content ...'/>
```



- OU Podcasts (data.open.ac.uk)

Great-circle distance

<http://data.open.ac.uk/podcast/218dce44a4ed17b36ada50d18b866b03>

<i><u>hasiTunesU</u></i>	<i>http://deimos.apple.com/WebObjects/Core.woa/Browse/itunes.open.ac.uk.21332</i>
<i><u>relatesToCourse</u></i>	<i>mu120</i>
<i><u>transcript</u></i>	<i>mu120.04showing04.pdf</i>
<i><u>depiction</u></i>	<i>mu120-showing-the-way_00359_std.jpg</i>
<i><u>download</u></i>	<i>mu120.04showing04.m4v</i>
<i><u>duration</u></i>	<i>00:03:39</i>
<i><u>description</u></i>	<i>Great-circle distances might be the shortest way to travel, but they are not always easiest. We also look at how lines of latitude aren't great-circles and that permits must be acquired to fly over many countries.</i>
<i><u>isPart</u></i>	<i>58dbd5be4f01f4b1eec1df1e8f97eaad</i>
<i><u>published</u></i>	<i>2009-05-19T02:29:55+01:00</i>



- Dublin Core
 - <http://dublincore.org/documents/dcmi-terms/> (**dcterms**)
- FOAF
 - <http://xmlns.com/foaf/spec/> (**foaf**)
- The W3C ontology for media resources
 - <http://www.w3.org/TR/mediaont-10/> (**ma**)
- The Media Vocabulary
 - <http://payswarm.com/vocabs/media> (**media**)
- The Nice Tag Ontology
 - <http://ns.inria.fr/nicetag/2010/09/09/voc.html> (**nt**)

- Main design decisions:
 1. The selected base URI is:
<http://linkeduniversities.org>
 2. VideoLectures objects are represented as `media:Recording`, authors are represented as `foaf:Person`
 3. The video title is duplicated in the properties `rdfs:label` & `dcterms:title`

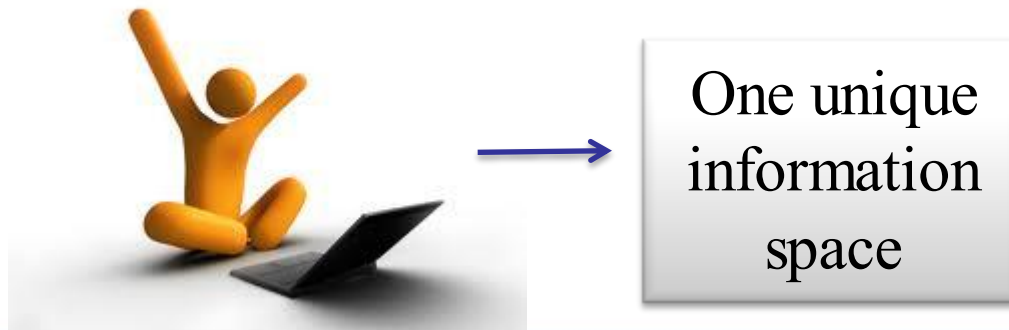
4. The set of tags and categories associated to a video is represented by the `nt:isRelatedTo` property
5. The assigned classification in the unified search space is represented by the `dcterms:subject` property



<http://linkeduniversities.org/video/CarnegieMellonU/youtube/B135229F3706D215>

rdf:type	media:Recording
media:download	http://www.youtube.com/watch?v=TOTuStPIeFc&feature=youtube_gdata_player
dcterms:title	CMU Football Engineering Summer 2008 Video
rdfs:label	CMU Football Engineering Summer 2008 Video
dcterms:description	Football [...]Summer 2008 Video
foaf:thumbnail	http://i.ytimg.com/vi/TOTuStPIeFc/3.jpg
media:duration	155
dcterms:isPart	http://linkeduniversities.org/video/CarnegieMellonU/youtube/playlist/B135229F37
ma:publisher	http://linkeduniversities.org/video/CarnegieMellonU/youtube/user/footballtracking
dcterms:published	2011-06-03T23:23:53.262Z
nt:isRelatedTo	http://linkeduniversities.org/video/CarnegieMellonU/tag/sports
nt:isRelatedTo	http://linkeduniversities.org/video/CarnegieMellonU/tag/football
dcterms:subject	http://dmoz.org/Sports/Football/Rugby_Union
dcterms:subject	http://linkeduniversities.org/video/CarnegieMellonU/dmoz/Sports/Football/Rugby_Union

- Select a common categorization scheme, requirements:
 1. Be general, aim to cover all subjects in “the universe of information”
 2. Be fully public
 3. Be available in RDF



- Candidates:

1. The Open Directory Project (DMOZ) 

<http://www.dmoz.org/rdf.html>

2. DBpedia Categories

<http://dbpedia.org/About>

3. Library of Congress Subject Headings

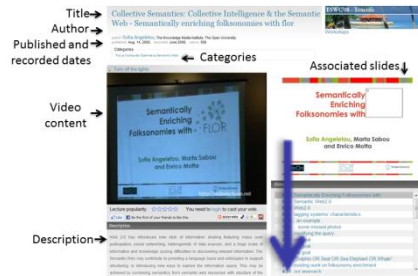
<http://id.loc.gov/authorities/about.html>

4. The International Press

Telecommunications Council (IPTC)

<http://www.iptc.org/site/NewsCodes/>

(1) Extract the information from the video lecture



(4) Obtain the ODP document classification

Reference/Knowledge_Management
(id=495), w=0.71



(2) Generate an HTML document

```
<html>
<head>
  <title>
</title>
  <meta name = keywords
    content= Computer Science, Semantic Web>
</head>
<body>
<p> Web 2.0 ... of the Semantic Web </p>
</body>
</html>
```



(3) Provide the document to the textwise classification service



- Step (2) Generate an HTML document
 - Video title = HTML title
 - Video description = HTML body
 - Video tags & source-dependent categories = HTML meta keywords

```
<html>
  <head>
    <title> Collective Intelligence [...] folksonomies </title>
    <meta name = keywords content= Computer Science, Semantic Web>
  </head>
  <body> <p>Web 2.0 has introduced [...] the Semantic Web</p>
</body>
</html>
```

- Evaluation measures:
 - Coverage
 - How many video lectures have been assigned at least one ODP category?
 - Correctness
 - Which percentage of the assigned categories are considered correct?
 - Specialization
 - Is it possible to find a more refined ODP category to describe the same video content?

- Evaluation set up:
 - 25 randomly selected videos for each source
 - 3 engaged evaluators
 - Task: evaluate the classification of each video
 - 0 -> incorrect
 - 1 -> correct but not specialized
 - 2 -> correct and specialized
 - Information provided to perform the task:
 - The information extracted from the video
 - The ODP categories assigned to the video
 - The complete ODP hierarchical classification



- New LD educational dataset
 - More than **14,000** video lectures from **27** different institutions sharing the same search/browsing space
 - Quality
 - Coverage: **98%** of the videos were assigned at least one ODP category
 - Correctness and Specialization:
 - **89%** correct classifications
 - **51%** specialized classifications
 - Fleiss' kappa statistic $k=0.71$

- Lessons Learned

- LD is simple, getting data and remodelling it is a high-cost process
- Need to agree on a set of vocabularies
- Need to agree on common searchable space
- Need to establish qualitative criteria and quantitative evaluation measures
- Educational LD is not about a killer app, but about multiple small things that are made easier!



<http://smartproducts1.kmi.open.ac.uk/web-linkeduniversities/index.htm>

linkeduniversities.org/video Dataset

Description

Linked Universities is an alliance of european universities engaged into exposing their public data as linked data. Linked data is a set of principles to put raw data on the Web, making them Web addressable and linkable, so that they can be easily accessed, discovered, connected and reused. The idea is that data from different institutions and organisations can contribute to a common data space on the Web: the Web of Data.

There are only a few universities currently exposing their public data as linked data, using technologies such as RDF and SPARQL to give direct access to information such as their publications, courses, educational material, etc. In addition, initiatives are currently often disconnected from each other. For every new site being developed, a lot

[Description](#)

[Documentation](#)

[Data](#)

[Contact](#)