Semantic technologies for Web information integration

Stefan Dietze
Introduction

- **L3S Research Center** (Hannover, Germany)

  ⇒ [http://www.l3s.de/](http://www.l3s.de/)

- **Knowledge Media Institute** (KMI) of The Open University (Milton Keynes, UK)

  ⇒ [http://kmi.open.ac.uk/](http://kmi.open.ac.uk/)
Research interests

- Semantic Web & Linked Data
- Data & knowledge integration: mapping, classification, interlinking, enrichment
- Application domains: education/TEL & Web archiving

Projects & activities

- EU funded research projects:
  - Meducator
  - Meme
  - NoTube
  - LinkedUp
  - Luisa

- (Linked) Web data & education
  - „Linked Learning“ and „LALD“ workshop series (eg LILE2012@WWW2012)
  - http://linkededucation.org

⇒ http://purl.org/dietze
Web information integration

Challenges

- Heterogeneity and lack of interoperability due to diversity of content and data:
  - social media, static Web sites & documents
  - structured data and metadata (RDBMS, XML, RDF, ...)
  - Web APIs and services (REST, SOAP, OAI-PMH)
- Extraction of structured data (from unstructured resources such as text, images and video)
- Exposing & integrating data on the Web
- Data interlinking and correlation
(Linked) Open Data

wider movement towards ‘Open’ Data

“Stop hugging your data…”
Sir Tim Berners-Lee, 2009

TED Ideas worth spreading
(Linked) Open Data

Wider movement towards ‘Open’ Data

"Stop hugging your data…”
Sir Tim Berners-Lee, 2009

Linked (Open) Data – “Semantic Web done right”

- W3C standards (RDF, SPARQL) to expose data, URIs to interlink datasets
- Vision: well connected graph of open Web data
- => vast cloud of interconnected datasets (http://ckan.net/group/lodcloud, 300+ datasets, 30+ billions of triples)
(Linked) Open Data

wider movement towards ‘Open’ Data

“Stop hugging your data…”
Sir Tim Berners-Lee, 2009

educator
world wide med

SmartLink

Stefan Dietze 15/05/12
(Linked) Open Data

wider movement towards ‘Open’ Data

“Stop hugging your data…”
Sir Tim Berners-Lee, 2009

educator
world wide med

SmartLink

arcomem
ARCOMEM

- EC-funded Integrating Project (IP)
- Started in February 2011 (3 years duration)

- 12 partners:

=> http://www.arcomem.eu
ARCOMEM: towards focused Web archiving

Challenges

- Vast and **highly dynamic** Web content:
  “After 1 year only 40% of Web pages are still accessible while 60% of the pages are new” [1]

- **Towards Web Archiving**: „Archive and Store Everything“ is not practical

ARCOMEM: towards focused Web archiving

Challenges

- Vast and highly dynamic Web content:
  “After 1 year only 40% of Web pages are still accessible while 60% of the pages are new” [1]

- Towards Web Archiving: „Archive and Store Everything“ is not practical

Vision/goal:

- Towards highly focused, community archives
  => Focused Web crawling and archiving and
  => Semantic search & navigation of Web archives via ...
  => ...extraction of “entities”, “topics”, “opinions”, “events” (ETOE)

ARCOMEM approach & outcomes

Approach

1. Exploit the **Social Web** to bootstrap ETOE by detecting
   - Events and entities (e.g. sport events, elections, etc)
   - Topics (e.g. global warming, financial crisis, swine flu) & opinions

2. Exploit the **Semantic Web**/(Linked) Web Data to enrich, disambiguate and correlate extracted data (=ETOE)

Outcomes

⇒ **Focused Web archives** (ARCOMEM use cases)
⇒ Comprehensive **dataset/knowledge base (LOD) of Web content metadata** which allows to navigate through Web content and its evolution
⇒ **Tools and methods** for focused crawling and ETOE detection
ARCOMEM use cases: media and political archives

Two exemplary prototypical applications

- Media-related Web archives
  => broadcasters SWR & Deutsche Welle

- Social-web-aware political Web Archives
  => Hellenic and Austrian Parliaments

Advantages

- Shaping ARCOMEM technology based on particular application needs
- Requirements and data for focused crawling and analysis (e.g. Rock am Ring SWR content, financial crisis data from parliaments etc)
- Verification of research results in real world contexts
ETOE extraction from Web objects
**Research challenges**

- **Entity recognition** from degraded input sources (tweets etc)
  - Advancing state of the art NLP and text mining
  - Dynamics detection: evolution of terms/entities

- **Semantic representation** of Web objects and entities
  - Appropriate RDF schemas for ETOE and Web objects
  - Exploiting (Linked Open) Web data to enrich extracted ETOE

- **Entity classification** (into events, locations, topics etc) & **consolidation**
Processing cycle

- **Processing and entity extraction** from sets of crawled Web objects
- Interaction with storage to retrieve objects and **write entities into knowledge base** (OpenRDF/OWLIM, distributed Hbase RDF store)
RDF schema for ARCOMEM knowledge base

- Relationships between ARCOMEM entities (ETOE etc) and information objects
- RDF schema: http://www.gate.ac.uk/ns/ontologies/arcomem-data-model.rdf
(ETOE) extraction from text => GATE

- ETOE extraction from text via novel components for GATE (Sheffield, http://gate.ac.uk/)
- Text mining, NLP, entity recognition and classification

Logic Layer

Gate 1: Preprocessing & Entity Extraction
  - Language Detection
  - Linguistic Preprocessing
  - Entity Extraction
  - Video Processing
  - Picture Processing
  - Picture Analysis

Gate 2: Events Extraction
  - Relation Extraction
  - Event Extraction
  - Opinion Mining
  - Topic Extraction

Gate 3: Opinion Mining

Social Web Analysis
Social Web Analysis (WP2)
Linked Data Enrichment
(RDF) Data Validation/Storage

Storage Layer

ARCOMEM TEMP Storage
- ARCOMEM TEMP DOCUMENT/OBJECT STORE
- ARCOMEM TEMP KB

ARCOMEM LIVE Storage
- ARCOMEM DOCUMENT/OBJECT STORE
- ARCOMEM LIVE KB

Stefan Dietze 15/05/12
ARCOMEM research challenges:

- Text processing in **multiple languages** (automated language detection)
- Language processing & entity recognition on **social media/degraded texts** (e.g. tweets)
- Entity classification (particularly wrt ETOE)

Progress so far:

- 3 components for (a) **term recognition**, (b) **entity recognition**, and (c) **event detection**
- Languages: English & German (automated language detection)
- Applied to ARCOMEM use case data:
  - **Greek financial crisis dataset**: 84 Web documents from news sites, forums etc
  - **SWR Rock am Ring festival**: 51 HTML documents (>3000 user comments)
  - **Austrian Parliament crawl**: ca 250 HTML documents
The Greek Crisis: Trichet rejects ECB role as lender of last resort

Tuesday, October 4, 2011
Trichet rejects ECB role as lender of last resort
Financial Times
October 4, 2011

Jean-Claude Trichet has dashed hopes that the European Central Bank will ride to the rescue of the eurozone by pledging to backstop crisis-hit member states.

In one of his last appearances as ECB president, Mr Trichet rejected the idea of the ECB acting as lender of last resort to governments. It was up to eurozone political leaders to restore investor confidence in Europe’s monetary union, he told the European Parliament.

candidate multi-word term
ETOE extraction results so far

- **Example entities (types):**
  - ECB (Organisation),
  - Athens (Location),
  - Jean Claude Trichet (Person)

- **Example queries:**
  1. Simple: Get **Web Objects** about **events** of type "industrial action"
     => [http://tinyurl.com/78ny7p5](http://tinyurl.com/78ny7p5)
  2. Correlated: Get **Web objects** about **events** (arco:Event) in **Athens** (arco:Location) (involving the **IMF** (arco:Organisation))
     => [http://tinyurl.com/78uj5at](http://tinyurl.com/78uj5at)

<table>
<thead>
<tr>
<th>Type</th>
<th>#Entities</th>
</tr>
</thead>
<tbody>
<tr>
<td>arco:Time</td>
<td>51416</td>
</tr>
<tr>
<td>arco:Money</td>
<td>6335</td>
</tr>
<tr>
<td>arco:Event</td>
<td>759</td>
</tr>
<tr>
<td>arco:Organisation</td>
<td>15376</td>
</tr>
<tr>
<td>arco:Location</td>
<td>21218</td>
</tr>
<tr>
<td>arco:Person</td>
<td>4465</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>99569</strong></td>
</tr>
</tbody>
</table>

(+ large number of terms)
ETOE extraction from images and video

- **University of Southampton:** *initial prototypes & experiments*
- **Results so far:** plenty low-level features, less ETOE
- **Workaround:** corpus of annotated media for similarity matching
Data consolidation and integration problem

- Data extracted from **different components** or during **different processing cycles** not aligned
  => **consolidation, disambiguation & correlation** required.

<Location>Greece</Location> <Person>Venizelos</Person>

<Location>Griechenland</Location> <Organisation>Greek Parliament</Organisation>
**Data clustering & enrichment**

- **Enrichment of entities** with related references to **Linked Data**, particularly **reference datasets** (DBpedia, Freebase, ...)
- **=>** use enrichments for **correlation/clustering/consolidation (L3S)**
Enrichment for clustering and correlation: example

<Person>Jean Claude Trichet</Person>

<Event>Trichet warns of systemic debt crisis</Event>

<Organisation>ECB</Organisation>
Enrichment for clustering and correlation: example

Jean Claude Trichet

Trichet warns of systemic debt crisis

ECB

http://dbpedia.org/resource/Jean-Claude_Trichet

http://dbpedia.org/resource/ECB
Enrichment for clustering and correlation: example

- **Event**: Trichet warns of systemic debt crisis
- **Person**: Jean Claude Trichet
- **Organisation**: ECB

Enrichment:

- [Jean-Claude Trichet](http://dbpedia.org/resource/Jean-Claude_Trichet)
- [ECB](http://dbpedia.org/resource/ECB)

Properties:

- `dbpprop:office`:
  - `dbpedia:President_of_the_European_Central_Bank`
  - `dbpedia:Governor_of_the_Banque_de_France`
- `dcterm:subject`:
  - `category:Living_people`
  - `category:Karlspreis_recipients`
  - `category:Alumni_of_the_École_Nationale_d'Administration`
  - `category:People_from_Lyon...`
ARCOMEM entities and enrichments - graph

- Nodes: entities/events (blue), enrichments DBpedia (green), Freebase (orange)
- 1013 clusters of correlated entities/events
ARCOMEM entities and enrichments - graph

- Nodes: entities/events (blue), enrichments DBpedia (green), Freebase (orange)
- 1013 clusters of correlated entities/events => cluster expansion by considering related enrichments
Clustering of entities via enrichment relatedness

Discovery of “related” entities by discovering related enrichments

(a) Retrieving possible paths between 2 enrichments (eg via RelFinder
http://www.visualdataweb.org/relfinder.php)

(b) Computation of relatedness measure (considering variables such as shortest path, number of paths, relationship types, number of directly connected edges of both enrichments...)

(c) Clustering enrichments (entities) which are above certain threshold
### Enrichment evaluation results

- Manual evaluation of 240 enrichment-entity pairs
- Available scores: 1 (correct), 0 (incorrect), 0.5 (vague or ambiguous relationship)

<table>
<thead>
<tr>
<th>Entity Type</th>
<th>Average score DBPedia</th>
<th>Average score Freebase</th>
<th>Average Score Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>arco:Event</td>
<td>0.71</td>
<td></td>
<td>0.71</td>
</tr>
<tr>
<td>arco:Location</td>
<td></td>
<td>0.94</td>
<td>0.88</td>
</tr>
<tr>
<td>arco:Money</td>
<td>0.67</td>
<td></td>
<td>0.67</td>
</tr>
<tr>
<td>arco:Organization</td>
<td>0.93</td>
<td>1</td>
<td>0.97</td>
</tr>
<tr>
<td>arco:Person</td>
<td>0.9</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td>arco:Time</td>
<td>0.74</td>
<td></td>
<td>0.74</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>0.79</strong></td>
<td><strong>0.94</strong></td>
<td><strong>0.87</strong></td>
</tr>
</tbody>
</table>
Outlook

Short term

- Investigation of reasons for enrichment noise
  - Ambiguous entities with no context (e.g. Athens in Greece vs. Athens in Greene County, New York).
  - Flaws in DBpedia Spotlight results, e.g. “Greek strategy on debt crisis” vs. “strategy games”
- Data quality in general
- ARCOMEM workshop (at ISWC2012?) on knowledge extraction and consolidation (from social media)

Longer term

- Publication of ARCOMEM ETOE dataset
- Release of ETOE detection and clustering methods as general purpose tools
"Stop hugging your data..."
Sir Tim Berners-Lee, 2009
mEducator

- EC-funded eContentPlus Best Practice Network (BPN)
- May 2009 – May 2012 (3 years duration)

- 14 partners:

=> http://www.meducator.net
Motivation: educational resources/data on the Web

State

- Vast Open Educational Resource (OER) metadata collections
  (e.g. MIT OpenCourseware, ARIADNE, OpenLearn)
- Usually exposed via APIs/services
- Competing Web interfaces
  (e.g. SQI, OAI-PMH, SOAP, REST)
- Competing metadata standards
  (e.g. IEEE LOM, ADL SCORM, DC...)
- Competing exchange formats and serialisations
  (e.g. JSON, RDF, XML)

Issues

- Heterogeneity & lack of interoperability
- Lack of take-up
Motivation: challenges

1. Improving **OER metadata interoperability** by applying LOD principles...

2. ...while exploiting **existing OER metadata and infrastructures**

![Diagram showing Open Educational Resources and Linked Data](image-url)
LD for integration of educational services and data

Educational application & presentation

Educational data & services integration

Linked Education

API layer

Educational Services Linked Data

Educational Resources Linked Data

Educational Web data & services

I. Educational services integration

II. Educational data integration

Stefan Dietze 15/05/12
Application context: biomedical education

Metamorphosis+ => http://metamorphosis.med.duth.gr/

Tailored (L)CMS plugins => http://www.meducator3.net/

Stefan Dietze 15/05/12
General approach: educational services integration
SmartLink for services annotation & discovery

- **Linked Data** about (educational/scientific) **data repositories, their services and APIs**
SmartLink for services annotation & discovery

- **Linked Data** about (educational/scientific) **data repositories, their services and APIs**

- **Semantic search & execution of educational services** via dedicated API: data retrieval, **transformation/lifting** (eg, get resource metadata in RDF from educational, English stores in biomedicine)

  ⇒  [http://ckan.net/package/smartlink](http://ckan.net/package/smartlink)
  ⇒  [http://purl.org/smartlink](http://purl.org/smartlink)
SmartLink: data lifting into RDF (via lifting templates)

Comparison of written reports of mammography, sonography and magnetic resonance mammography

Patients with abnormal breast findings (n = 413) were examined by mammography, sonography and histologically. Sensitivity for mammography was 83.7%, for sonography it was 85.1% and for MR mammography and sonography in 26.2%, and by MR mammography in 66.7%. In nine patients 78.9% and by MR mammography in 68.4% of patients. Combination of all three diagnostic methods sonography combined was identical to that of MR mammography (ie 94.6%).
Educational data integration based on Linked Data

Educational application & presentation

Educational data & services integration

Educational services Linked Data

Linked Educational Resources

I. Educational services integration

II. Educational data integration

Web API

Educational resources

SmartLink

Educational Resources Linked Data

Mediator

Stefan Dietze 15/05/12
Data integration: enrichment, interlinking, clustering

3-step approach to enrich and interlink resulting RDF:

1) **Enrichment** with references to LOD datasets
   - (DBpedia, Freebase, Bioportal)

2) **Structural clustering** and correlation based on shared enrichments/metadata

3) **Linguistic clustering** and correlation based on metadata similarity
   - ⇒ [http://ckan.net/package/meducator](http://ckan.net/package/meducator)
   - ⇒ [http://linkededucation.org/meducator](http://linkededucation.org/meducator)
Data integration based on LD: (1) Enrichment

- **Automated** enrichment (DBpedia, Freebase) as built-in feature of RDF storage
- **Semi-automated** enrichment via Bioportal ([http://bioportal.bioontology.org/](http://bioportal.bioontology.org/))
(1) Enrichment: semi-automated via BioPortal

Metamorphosis+
http://metamorphosis.med.duth.gr/

user entered search term

related standardized terms from BioPortal

enriching properties ‘keyword’, ‘discipline’ and/or ‘specialty’ with chosen standardized term
(1) Enrichment: automated via DBpedia/Freebase

<table>
<thead>
<tr>
<th>Topic</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cervical cancer</td>
<td>59</td>
</tr>
<tr>
<td>Screening</td>
<td>31</td>
</tr>
<tr>
<td>Cervical</td>
<td>29</td>
</tr>
<tr>
<td>Hpv</td>
<td>29</td>
</tr>
<tr>
<td>Oxygenation</td>
<td>26</td>
</tr>
<tr>
<td>Childhood</td>
<td>22</td>
</tr>
<tr>
<td>differential_diagnosis</td>
<td>19</td>
</tr>
<tr>
<td>Knowledge</td>
<td>18</td>
</tr>
<tr>
<td>Learning</td>
<td>17</td>
</tr>
<tr>
<td>decision_making</td>
<td>16</td>
</tr>
<tr>
<td>Training</td>
<td>15</td>
</tr>
<tr>
<td>Lecture</td>
<td>15</td>
</tr>
<tr>
<td>Risk</td>
<td>15</td>
</tr>
<tr>
<td>hpv_infection</td>
<td>15</td>
</tr>
<tr>
<td>Fear</td>
<td>15</td>
</tr>
<tr>
<td>pap_smear</td>
<td>15</td>
</tr>
<tr>
<td>Abnormal</td>
<td>14</td>
</tr>
<tr>
<td>Ventilation</td>
<td>14</td>
</tr>
<tr>
<td>Ecg</td>
<td>14</td>
</tr>
</tbody>
</table>
(1) Enrichment: automated via DBpedia/Freebase

Frequency of DBpedia references in mEducator dataset (number of resources referring to enrichment)

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cervical_cancer</td>
<td>59</td>
</tr>
<tr>
<td>Screening</td>
<td>31</td>
</tr>
<tr>
<td>Cervical</td>
<td>29</td>
</tr>
<tr>
<td>Hpv</td>
<td>29</td>
</tr>
<tr>
<td>Oxygenation</td>
<td>26</td>
</tr>
<tr>
<td>Childhood</td>
<td>22</td>
</tr>
<tr>
<td>differential_diagnosis</td>
<td>19</td>
</tr>
<tr>
<td>Knowledge</td>
<td>18</td>
</tr>
<tr>
<td>Learning</td>
<td>17</td>
</tr>
<tr>
<td>decision_making</td>
<td>16</td>
</tr>
<tr>
<td>Training</td>
<td>15</td>
</tr>
<tr>
<td>Lecture</td>
<td>15</td>
</tr>
<tr>
<td>Risk</td>
<td>15</td>
</tr>
<tr>
<td>hpv_infection</td>
<td>15</td>
</tr>
<tr>
<td>Fear</td>
<td>15</td>
</tr>
<tr>
<td>pap_smear</td>
<td>15</td>
</tr>
<tr>
<td>Abnormal</td>
<td>14</td>
</tr>
<tr>
<td>Ventilation</td>
<td>14</td>
</tr>
<tr>
<td>Ecg</td>
<td>14</td>
</tr>
</tbody>
</table>

Clustering of mEducator graph (blue nodes: resources, green nodes: enrichments)
(1) Enrichment: evaluation

Evaluation categories:

<table>
<thead>
<tr>
<th>Category</th>
<th>Example source text</th>
<th>Enrichment example</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) – perfect match</td>
<td>“... to educate their patients about <strong>cervical cancer</strong> causation, screening and prevention.”</td>
<td><a href="http://dbpedia.org/resource/Cervical_cancer">http://dbpedia.org/resource/Cervical_cancer</a></td>
</tr>
<tr>
<td>(B) – broad match</td>
<td>“... in safety critical and restrictive environments, such as in aviation, <strong>space</strong>, automobile and heavy industrial machinery operation”</td>
<td><a href="http://dbpedia.org/resource/Space">http://dbpedia.org/resource/Space</a></td>
</tr>
<tr>
<td>(C) – no match</td>
<td>“EASA started its work in the <strong>fields</strong> of Airworthiness and will soon its competencies...”</td>
<td><a href="http://dbpedia.org/resource/Fields_Medal">http://dbpedia.org/resource/Fields_Medal</a></td>
</tr>
</tbody>
</table>

Evaluation results (sample size = 200):

<table>
<thead>
<tr>
<th>Category</th>
<th>n# enrichments</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>184</td>
<td>92</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>8</td>
<td>4</td>
</tr>
</tbody>
</table>
(3) Clustering (similarity-based, linguistic)

Vector-based similarity computation based on:

1) **Data indexing** => Doc-Term Matrix (term frequencies in given resource metadata)

2) **Creation of similarity matrices** => similarity values between resources

3) **Clustering** (based on similarity thresholds)
Exploratory search enabled via clustering (1/2)

Metamorphosis+
http://metamorphosis.med.duth.gr/

Search Results

- Boston Criteria for the Diagnosis of Congestive Heart Failure
- Complete heart block
- Complete heart block Repurposed
- ECCA TEST 3 Heart diseases
- Heart Anatomy
- Heart Structure
- Heart disease
- Powerpoint presentation on Heart disease (test)
- The human heart and it’s function - test

Find additional results by clicking the checkbox of the items above you are most interested in.
Exploratory search enabled via clustering (2/2)

Metamorphosis+
http://metamorphosis.med.duth.gr/

<table>
<thead>
<tr>
<th>Search Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston Criteria for the Diagnosis of Congestive Heart Failure</td>
</tr>
<tr>
<td>Complete heart block</td>
</tr>
<tr>
<td>Complete heart block Repurposed</td>
</tr>
<tr>
<td>ECCA TEST 3 Heart diseases</td>
</tr>
<tr>
<td>Heart Anatomy</td>
</tr>
<tr>
<td>Heart Structure</td>
</tr>
<tr>
<td>Heart disease</td>
</tr>
<tr>
<td>Powerpoint presentation on Heart disease (test)</td>
</tr>
<tr>
<td>The human heart and it’s function - test</td>
</tr>
</tbody>
</table>

Find additional results by clicking the checkbox of the items above you are most interested in.

Possibly related resources

<table>
<thead>
<tr>
<th>First order associations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete heart block Repurposed</td>
</tr>
<tr>
<td>Pericarditis</td>
</tr>
<tr>
<td>Pericarditis Repurposed</td>
</tr>
<tr>
<td>Right bundle branch block 2</td>
</tr>
<tr>
<td>Right bundle branch block 1</td>
</tr>
<tr>
<td>Ischaemic chest pain</td>
</tr>
<tr>
<td>Right bundle branch block 1 Repurposed</td>
</tr>
<tr>
<td>Right bundle branch block 2 Repurposed</td>
</tr>
<tr>
<td>Ischaemic chest pain Repurposed</td>
</tr>
<tr>
<td>Progressing myocardial infarction Repurposed</td>
</tr>
</tbody>
</table>

Second order associations
Data so far: SmartLink/mEducator in LOD cloud

- **SmartLink**
  - [http://ckan.net/package/smartlink](http://ckan.net/package/smartlink)
  - > 2000 triples so far
  - > 300 links to iServe
  - APIs used by several applications

- **mEducator**
  - [http://ckan.net/package/meducator](http://ckan.net/package/meducator)
  - > 35000 triples so far
  - > 1000 links to DBpedia & Bioportal ontologies
  - APIs used by 4 applications
Outlook

R&D goals

- Performance: caching of data (retrieved via SmartLink) instead of distributed queries
- Periodic data imports from remote repositories into new, large-scale repository at [http://data.linkededucation.org](http://data.linkededucation.org) (harvesting, transformation, enrichment, clustering)
- Exploiting DBpedia/Freebase graph to uncover indirect resource relationships

Coverage and outreach

- Integration of additional datasets => [http://linkededucation.org](http://linkededucation.org) => long-term plan: joint endpoint for all datasets
- Linking Web Data for Education Challenge
Outlook: “LinkedUp” Support Action

Linking Web Data for Education Project – Open Challenge in Web-scale Data Integration

- EC Support Action, start October 2012

Goals

- Push forward adoption of Web data/Linked Data in educational context
- Drive technological advancement of Web data integration technologies

Approach

- Open data competition (initial calls expected early 2013) incl. technical, legal and financial support
- Open Web education as big data scenario

Partners

+ network of associated institutions (eg Commonwealth of Learning, Athabasca University, Talis UK, ...)

Stefan Dietze
15/05/12
Thank you!

http://purl.org/dietze

http://linkededucation.org

http://www.l3s.de