ShareIt: Mining #SocialMedia Activities for Detecting #Events

Raphaël Troncy <raphael.troncy@eurecom.fr>
Cover of the December 25, 2006 issue
Quiz Test: who has already ...

1. edited a Wikipedia page?
2. shared photos on Flickr / Picassa?
3. uploaded a video on YouTube / Dailymotion?
4. used a mobile-aware application: Foursquare / Gowalla?
5. published a thought / comment on a blog?
6. published its status on Twitter / Identi.ca / FriendFeed?
7. shared bookmarks on Del.icio.us / Faviki?
8. own a Facebook account and does all this?
What do you do for getting event info?

http://s3mr.eu/agenda/

This official event page does a very poor job to bring structured information (Sander Koelstra)

Agenda

Sunday 26 June 2011
17:00-19:00 - Registration desk open
19:00-21:00 - Reception and dinner at the Marmara Antalya Hotel

Monday 27 June 2011
08:30-09:30 - Registration desk open
09:00-09:30 - Opening
09:30-10:45 - Lecture 1: Image and Video Tagging in the Internet Era (Dr Xian-Sheng Hua)
10:45-11:00 - Coffee break
11:00-12:30 - Lecture 1 (Dr Xian-Sheng Hua)
12:30-14:00 - Lunch
14:00-14:20 - Introduction to the group discussions
14:20-16:20 - Group discussion 1
16:20-16:40 - Coffee break
16:40-17:00 - Short presentation of group discussion results

Tuesday 28 June 2011
09:00-10:45 - Lecture 2: Human Sensing, Implicit Tagging, and Implicit Interaction (Prof Anton Nijholt)
10:30-10:45 - Coffee break
11:00-12:30 - Lecture 2 (Prof Anton Nijholt)
12:30-13:30 - Lunch
13:30-13:40 - Setup Poster session 1
13:40-14:00 - Poster session 1
14:00-14:50 - Setup Poster session 2
14:50-15:50 - Poster session 2
15:50-16:10 - Coffee break
16:10-17:55 - Group discussion 2
18:00-18:20 - Short presentation of group discussion results

Wednesday 29 June 2011
09:00-10:30 - Lecture 3: ShareIt: Mining SocialMedia Activities for Detecting Events (Dr Raphael Troncy)
Looking for more structured information?

INTRODUCTION
Multimedia content has become ubiquitous on the web, creating new challenges for indexing, access, search and retrieval. At the same time, much of this content is now available on social sharing websites like YouTube or Flickr, or shared on social networks like Facebook. In such environments, the content is usually accompanied with metadata, tags, ratings, comments, information about the uploader and their social network, etc.

Analysis of this "social media" shows a great...

Show more...
Looking for some media?
Looking for some media?

http://www.bing.com/images/search?q=3mr%202011
Anything on Flickr / YouTube?

Video Lectures reports 1 event and 3 lectures
SSMS participants were better “sharer”
Looking for some live information?

Not that much of activity on Twitter
Facebook is the place to be, right?
Music
Concerts
Dance
Sports
Weddings
Galleries
Exhibitions
Movies
Technology
Spirituality
Politics
Nightlife
Conferences
Networking
Science
Art
Food
Party
Theatre
Show
Museums
Lit.
Festivals
Acoustic
Rock
Learning
Kids
Campus
Health
Dinner
Birthdays
We have directory of events...
London PO, Eschenbach / Dvorak

When
Jan 2, 2010 7:30 pm (Saturday)

Where
Shanghai Oriental Arts Center (map)
Shanghai, 31
China

What

More details
 Improve this event listing

Twitter Hashtag: #EVlondo0102
London PO, Eschenbach / Dvorak

When
Jan 2, 2010 7:30 pm (Saturday)

Where
Shanghai Oriental Arts Center (map)
Shanghai, 31
China

What

Add to favorites
Save to calendar
Get countdown widget

More details

Improve this event listing

Twitter Hashtag: #EVLondo0102
Oriental Art Centre

by Verena Adlfinger

This photo is selected for Google Earth [?] - ID: 6681312
London PO, Eschenbach / Dvorak

When
Jan 2, 2010 7:30 pm (Saturday)

Where
Shanghai Oriental Arts Center (map)
Shanghai, 31
China

What

More details

Improve this event listing

Twitter Hashtag: #EVlondo0102
London Philharmonic Orchestra

The London Philharmonic Orchestra (LPO), based in London, is one of the major orchestras of the United Kingdom, and is based in the Royal Festival Hall. In addition, the LPO is the main resident orchestra of the Glyndebourne Festival Opera. The LPO also performs concerts at the Congress Theatre, Eastbourne and the Brighton Dome. The orchestra was formed in 1932 by Sir Thomas Beecham, and played its first concert on 7 October 1932 at the Queen’s Hall.

Read article at Wikipedia
Musical Career Began: London
Musical Genres: European Classical music
London PO, Eschenbach / Dvorak

When
Jan 2, 2010 7:30 pm (Saturday)

Where
Shanghai Oriental Arts Center (map)
Shanghai, 31
China

What

More details

Improve this event listing

Twitter Hashtag: #Evlondo0102
Antonín Leopold Dvořák was a Czech composer of Romantic music, who employed the idioms and melodies of the folk music of Moravia and his native Bohemia. His works include operas, symphonic, choral and chamber music. His best-known works include his New World Symphony, the Slavonic Dances, "American" String Quartet, and Cello Concerto in B minor.

**Property** | **Value**
--- | ---
$dbpedia-owl:Person/individualisedPnd$ | 11852836X
$dbpedia-owl:thumbnail$ | ![Antonín Leopold Dvořák](http://upload.wikimedia.org/wikipedia/commons/thumb/f/fb/Dvorak1.jpg/200px-Dvorak1.jpg)
dbpprop:abstract | 
- Antonín Leopold Dvořák was a Czech composer of Romantic music, who employed the idioms and melodies of the folk music of Moravia and his native Bohemia. His works include operas, symphonic, choral and chamber music. His best-known works include his New World Symphony, the Slavonic Dances, "American" String Quartet, and Cello Concerto in B minor.
- Antonín Leopold Dvořák [ˈantɔɲ̊iːn ˈleoːpɔlt ˈdvȏɾaːk] était un compositeur tchèque, né à Nelahozeves à 40 kilomètres au nord de Prague en Bohême, le 8 septembre 1841, et mort à Prague le 1 mai 1904. Son prénom est parfois orthographié « Anton », dans une forme germanisée, peu utilisée en langue française.
- Dvořák
There’s a lot of information out there…
A lot of information...
EventMedia Goals

1. Discover PAST, PRESENT and FUTURE events
2. Live, relive and predict experiences through shared media
3. Identify meaningful and/or interesting relationships between events/media/people
Agenda

- A crash course in the world of structured data
  - #microdata, #microformat, #rdfa
  - #rdf, #owl, #skos, #sparql, #linkeddata

- EventMedia (User-centered design approach)
  - LODE: a model for representing events
  - Scraping and interlinking description of events
  - Enriching events with illustrating media
  - Detecting events from social media activities

- Detecting events from human sensing
  - #twitter, #foursquare, #facebook
Fundamental shift:

From sending bits from one host to the other towards making sense of those bits
BelgianChocolates.com

Pralinés Deluxe Mix

2,99€/100g

Shopping Cart
From the Web to the Web of Data

Merchant Name

BelgianChocolates.com

Pralinés Deluxe Mix

2,99€/100g

Shopping Cart

Price

Product Name

Product Image
From the Web to the Web of Data

- How can website owners help Google make sense of their bits?

- Mark up their content using any of the following syntaxes:
  - Microdata
  - Micro format
  - RDFa

- "[...] We realized that structured data on the Web can and should accommodate multiple encodings."
domain-dedicated microformats

for specific, common, concise data

VS.

generic RDFa or embedded RDF

for custom data, RDF data, multiple schemas
About Microformats

Designed for humans first and machines second, microformats are a set of simple, open data formats built upon existing and widely adopted standards. Instead of throwing away what works today, microformats intend to solve simpler problems first by adapting to current behaviors and usage patterns (e.g. XHTML, blogging).
RDFa = a domain-independent way to explicitly embed your data
RDFa = a domain-independent way to explicitly embed RDF data
RDFa stands for…

RDF… in HTML … attributes
RDFa in attributes of a web page to...

... transfer data from an application to another through the web.

... write data only once for web users and web applications.
weaving RDFa in web pages
RDFa step 1
declare the schemas you are using
RDFa step 2

use attributes to mark, type and add data
RDFa step 3
let RDFa agents extract RDF from the document
Weekend off in Iona: 2006-10-21 to 2006-10-23. See Free time web site for info on Iona, UK.
Weekend off in Iona: 2006-10-21 to 2006-10-23. See Free time web site for info on Iona, UK.
schemas for data in this web page

<html xmlns="http://www.w3.org/1999/xhtml"
xmlns:cal="http://www.w3.org/2002/12/cal/icaltzd#"
xmlns:xs="http://www.w3.org/2001/XMLSchema#" >
<body>
<p about="#event1" instanceof="cal:Vevent">
  <b property="cal:summary">Weekend off in Iona</b>: 
  <span property="cal:dtstart" datatype="xs:date">2006-10-21</span> to 
  <span property="cal:dtend" datatype="xs:date">2006-10-23</span>. 
  see <a rel="cal:url" href="http://freetime.example.org/">Free time web site</a> for info on 
  <span property="cal:location">Iona, UK</span>.
</p>
</body>
</html>
Weekend off in Iona: 2006-10-21 to 2006-10-23.

see Free time web site for info on Iona, UK.
Web page data for an RDFa agent in this web page

<html xmlns="http://www.w3.org/1999/xhtml"
xmlns:calendar="http://www.w3.org/2002/12/cal/icaltzd#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#">
<body>
<p about="#event1" instanceof="calendar:Vevent">
<b property="calendar:summary">Weekend off in Iona</b>: 
<span property="calendar:dtstart" datatype="xsd:date">2006-10-21</span> to 
<span property="calendar:dtend" datatype="xsd:date">2006-10-23</span>. 
see <a rel="calendar:url" href="http://freetime.example.org/">Free time web site</a> for info on 
<span property="calendar:location">Iona, UK</span>.
</p>
</body>
</html>
data shared by both in this web page

Weekend off in Iona: 2006-10-21 to 2006-10-23. See Free time web site for info on Iona, UK.
what an RDFa agent knows from this web page

#event1 isA cal:Vevent
#event1 cal:summary "Weekend off in Iona"
#event1 cal:dtstart "2006-10-21"^^xs:date
#event1 cal:dtend "2006-10-23"^^ xs:date
#event1 cal:url <http://freetime.example.org/>
#event1 cal:location "Iona, UK"
RDF is the first layer of the Semantic Web standards.
RDF stands for Resource Description Framework
RDF is a triple model i.e. every piece of knowledge is broken down into

( subject , predicate , object )
image.jpg has for creator Raphael and depicts the elephant Ganesh
image.jpg has for creator Raphael
image.jpg depicts the elephant Ganesh
( image.jpg, creator, Raphael )
( image.jpg, depicts, Elephant Ganesh )

( subject, predicate, object )
in **RDF** the atoms of knowledge are triples of the form

\[(subject, predicate, object)\]
RDF triples can be seen as arcs of a graph (vertex, edge, vertex)
creator: Raphael

depicts: Ganesh

creator: Ganesh

depicts: Raphael

image.jpg
in **RDF** resources and properties are identified by URIs

http://mydomain.org/mypath/myresource
in RDF values of properties can also be literals i.e. strings of characters
http://www.cwi.nl/~troncy#me

http://purl.org/dc/elements/1.1#creator

http://flickr.com/photos/rtroncy/2923/

http://xmlns.com/foaf/0.1#depicts

Elephant Ganesh
The RDF Data Model

- An RDF document is an unordered collection of statements, each with a subject, predicate and object (aka triples)
- A triple can be thought of as a labelled arc in a graph
- Statements describe properties of web resources
- A resource is any object that can be pointed to by a URI:
  - a document, a picture, a paragraph on the Web, etc.
- Properties themselves are also resources (URIs)
Example of RDF Graphs

- HOUSE: http://example.com/damian_house
  - resident
  - address
    - number
    - street
  - Cranbrook Road
  - Bristol

- PERSON: http://example.com/damian
  - name
    - rdfs:seeAlso
      - http://example.com/document_b.rdf
  - Damian Steer

- Article1
  - has Author
    - CodyBurleson
      - has friend
        - MarySmith
          - author of
            - Article2
            - ArticleX
        - JohnDoe
          - author of
            - Article3
            - ArticleA
<div xmlns:v="http://rdf.data-vocabulary.org/#" typeof="v:Event">
  <a href="http://www.example.com/events/poisel_offenback.html" rel="v:url" property="v:summary">Philipp Poisel in Offenbach</a>
  <span property="v:description">See Philipp Poisel in Offenbach</span>
  When: <span property="v:startDate" content="2011-01-16T19:00-01:00">Jan 16, 7:00PM</span>
  <span property="v:endDate" content="2011-01-16T21:00-01:00">9:00PM</span>
  Where: <span rel="v:location"><span typeof="v:Organization">
    <span property="v:name">Capitol</span>,
    <span rel="v:address"><span typeof="v:Address">
      <span property="v:street-address">Kaiserstrae 106</span>,
      <span property="v:locality">Offenbach am Main</span>,
    </span></span>
    <span rel="v:geo"><span typeof="v:Geo">
      <span property="v:latitude" content="50.10945"></span>
      <span property="v:longitude" content="8.76579"></span>
    </span></span>
  </span></span>
  Category: <span property="v:eventType">Concert</span>
</div>
Rich Snippet Preview

Philipp Poisel at Tresor (Berlin) on 6 Sep 2010 – Last.fm

Sep 6, 2010 ... Last.fm concert page for Philipp Poisel at Tresor (Berlin) on 6 Sep 2010. Discuss the gig, get concert tickets, see who's attending, ...

Offenbach, Germany Sun, Jan 16, 2011
Münster, Germany Mon, Jan 17, 2011
Osnabrück, Germany Wed, Jan 19, 2011

www.last.fm/event/1657301+Philipp+Poisel - Cached
Rich Snippet Preview for Reviews

Xanh Restaurant
- 1 star to start. Food (1 star): I liked the food a lot here! The Xanh Salad is enormous and comes with some very tender beef...

Xanh Bistro
- 16161 Brookhurst St, Fountain Valley, CA 92708. Open 11am - 2:30pm. Dinner: 5pm - 10pm. Closed Tuesdays.

Xanh Outdoor
- 262 reviews. Price range: $$. XANH is a Vietnamese restaurant in Mountain View, CA. Read 262 reviews, view the menu and photos, and make reservations online for XANH.

- Rating: 7.2/10 - from 2,724 users. Directed by Anh Hung Tran. With Tran Nu Yến-Khé, Man San Lu, Thi Loc Truong. Mui goes to live with a pianist who has a fiancé. The pianist and Mui have a...
Rich Snippet Preview for People

Google search results for Pravir Gupta show his profile on Facebook, LinkedIn, and professional directories.
Rich Snippet Preview for Recipes

Google search results for "baked ziti" showing different recipe options and images related to baked ziti.
Rich Snippet Preview for Events

The Fillmore Concert Tickets, Schedule, Seating Chart | Official ... ☆
Get email alerts and never miss your favorite events at The Fillmore. Please enter your e-mail address. That is not a valid e-mail address format. ...
www.thefillmore.com/ - Cached - Similar

The Fillmore San Francisco - The Fillmore Schedule | Eventful ☆
View The Fillmore’s upcoming event schedule and profile - San Francisco, CA. The Fillmore, also known as Fillmore Auditorium, is located in San ...

Carolina Chocolate Drops Thu, Jun 23
Josh Ritter & the Royal City Band Thu, Jun 24
Robert Earl Keen Sat, Jun 26

eventful.com » San Francisco venues - Cached - Similar

Fillmore Events: Events in Fillmore, California ☆
Fillmore Events Directory. Includes listings for Events in Fillmore, California.
www.california-coast-worldweb.com/Fillmore/Events/ - Cached - Similar

San Francisco The Fillmore Events, Shows & Things to do - SF Gate ☆
Find 48 San Francisco The Fillmore events and show tickets and more on Zvents. Popular The Fillmore Events are Salsa Festival on the Fillmore, Fillmore Jazz ...
events.sfgate.com/san-francisco-ca/events/the+fillmore - Cached

New York Fillmore Events: Events, Shows & Things to do - NY Daily News ☆
Find 29 New York Fillmore Events events and show tickets and more on Zvents. Popular Fillmore Events Events are On Fillmore Plus Rachel Grimes, ...
events.nydailynews.com/new-york-ny/events/fillmore+events - Cached

Charlotte Charlotte Fillmore Events: Shows & Things to do - The ... ☆
Find 8 Charlotte Charlotte Fillmore events and show tickets and more on Zvents. Popular Charlotte Fillmore Events are Smashing Pumpkins, Adam Lambert with ...
events.charlotteobserver.com/charlotte-nc/events/charlotte+fillmore - Cached
Yahoo! Enhanced Results

Enhanced result with deep links, rating, address.
Harrison Ford - Wikipedia, the free encyclopedia
Early life  |  Early career  |  Milestone...  |  Other film...
Harrison Ford is an American film actor and producer. Ford is best known for his performances as Han Solo in the original Star Wars trilogy and as the title character of the Indiana... en.wikipedia.org/wiki/Harrison_Ford - Cached

Harrison Ford Web
Offers news, screen captures, promotional pictures, articles, interviews, multimedia clips,
Snippet generation using metadata

- Yahoo displays enriched search results for pages that contain microformat or RDFa markup using recognized ontologies
  - Displaying data, images, video
  - Example: GoodRelations for products
  - Enhanced results also appear for sites from which we extract information ourselves

- Also used for generating facets that can be used to restrict search results by object type
  - Example: “Shopping sites” facet for products

- Documentation and validator for developers
  - [http://developer.search.yahoo.com](http://developer.search.yahoo.com)

- Formerly: SearchMonkey allowed developers to customize the result presentation and create new ones for any object type
How search engines get this data?
Behind the scene

- With RDFa markup:
  
  ```html
  <div xmlns:v=http://rdf.data-vocabulary.org/#
       typeof="v:Review-aggregate">
    <span rel="v:itemreviewed">
      <h1 property="v:name">Drooling Dog Bar B Q</h1>
      <img rel="v:rating" src="stars_map.png" alt="4 star rating" />
      <em>based on <span property="v:count">15</span> reviews</em>
    </span>
  </div>
  ```

- With Micro-format markup:
  
  ```html
  <div class="hreview-aggregate">
    <span class="item vcard">
      <h1 class="fn org">Drooling Dog Bar B Q</h1>
      <img class="rating average" src="stars_map.png" alt="4 star rating" />
      <em>based on <span class="count">15</span> reviews</em>
    </span>
  </div>
  ```
Get your markup with test tool

Rich Snippets Testing Tool

Rich Snippets allows you to enhance your Google search results by marking up web pages with Microformats, RDFa or Microdata.

Test your website

Enter a web page URL to see how it may appear in search results:

http://www.urbanspoon.com/r/6/765421/restaurant/Pizza-My-Heart-

Examples: Urbanspoon, LinkedIn

Google search preview

Pizza My Heart - Sarta Cruz | Urbanspoon

10 reviews - Price range: Under $10 per entree
Excerpt from the page will show up here. Excerpt from the page will show up here.
Excerpt from the page will show up here. Excerpt from the page will show up here.
www.urbanspoon.com/r/6/765421/restaurant/Pizza-My-Heart-Santa-Cruz - Cached - Similar pages

Note that there is no guarantee that a Rich Snippet will be shown for this page on actual search results. For more details, see the FAQ

Extracted Rich Snippet data from the page

<hreview-aggregate
item hcard
fn = Pizza My Heart
org
organization-name = Pizza My Heart
adr
street-address = 1116 Pacific Ave Ste B
locality = Santa Cruz
region = CA
 postal-code = 95060
tel
value = (831) 426-2511
url = http://www.pizzamyheart.com/
How much structured data is out there?

One million web pages sampled from the Internet

<table>
<thead>
<tr>
<th></th>
<th>Microformats</th>
<th>RDFa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total pages</td>
<td>40,091</td>
<td>2,514</td>
</tr>
<tr>
<td>hCard / People</td>
<td>33,675 (13%)</td>
<td>1,160</td>
</tr>
<tr>
<td>Reviews</td>
<td>1,950 (88%)</td>
<td>872 (66%)</td>
</tr>
<tr>
<td>Recipe</td>
<td>152 (53%)</td>
<td>--</td>
</tr>
<tr>
<td>hCalendar / Event</td>
<td>126 (41%)</td>
<td>--</td>
</tr>
<tr>
<td>Products</td>
<td>519</td>
<td>77</td>
</tr>
</tbody>
</table>

Data from June 2010. Percentages: actually used data.
US/English Rich Snippets Usage

Searches on Google with rich results

2x since Oct 2009
Worldwide Rich Snippets Usage

Searches on Google with rich results

4x since Oct 2009
RDFa on the rise (Peter Mika@W3C Bilbao)

510% increase between March, 2009 and October, 2010

Percentage of URLs with embedded metadata in various formats
Future for Rich Snippets?

Philipp Poisel – Tour dates and concerts 2011
Find Philipp Poisel live concert tour dates, tickets, reviews, and more. Be the first to know when Philipp Poisel is playing live in your town!

Contacts also interested in Philipp Poisel:
Michael
Tom

Bremen (Kulturzentrum Schlachthof)
Sat, Feb. 12, 08:00pm
€24,10 via Eventim.de
€23,95 via TicketCenter.de
€24,90 via TicketOnline.com

Also going:
Ivan - Contact

Hamburg (Docks Hamburg)
Sun, Feb. 13, 08:00pm
€21,65 via Eventim.de
€24,90 via TicketOnline.com

www.philipp-poisel.de/termine/ - Cached - Similar

Even Richer Snippets: Using information form the user's social graph, given granted access; Direct price comparison.
Future for Rich Snippets?

Videos for pulp fiction robbery scene

Pulp Fiction Opening Restaurant Scene
3 min – Nov. 2010
Uploaded by qtarantino
www.youtube.com

In this Scene:
Amanda Plummer – IMDb
(Honey Bunny – Yolanda)

Tim Roth – IMDb
(Pumpkin – Ringo)

Yolanda: I love you, Pumpkin. – Play from here
Ringo: I love you, Honey Bunney. – Play from here

Fake mock-up. Authors' private view.

Even Richer Snippets using multimedia semantics.
What is Schema.org?

This site provides a collection of schemas, i.e., html tags, that webmasters can use to markup their pages in ways recognized by major search providers. Search engines including Bing, Google and Yahoo! rely on this markup to improve the display of search results, making it easier for people to find the right web pages.

Many sites are generated from structured data, which is often stored in databases. When this data is formatted into HTML, it becomes very difficult to recover the original structured data. Many applications, especially search engines, can benefit greatly from direct access to this structured data. On-page markup enables search engines to understand the information on web pages and provide richer search results in order to make it easier for users to find relevant information on the web. Markup can also enable new tools and applications that make use of the structure.

A shared markup vocabulary makes easier for webmasters to decide on a markup schema and get the maximum benefit for their efforts. So, in the spirit of sitemaps.org, Bing, Google and Yahoo! have come together to provide a shared collection of schemas that webmasters can use.

We invite you to get started!
What is Schema.RDFS.org?

Only very recently the three big search engines Bing, Google and Yahoo! introduced Schema.org, a collection of schemas that webmasters can use to markup their pages. The search engines rely on this markup to improve the display of search results, making it easier for people to find the right web pages.

This site is a complementary effort by people from the Linked Data community to express the terms provided by the Schema.org consortium in RDF. We currently provide static RDFS documents of the Schema.org terms in the formats listed below – and yes, we're heavily working on more ;) 

Schema.org types and properties in various formats (auto-updated daily):

- RDF/Turtle
- RDF/XML
- RDF/NTriples
- JSON
2011/06/27: [Announcement]

- We have posted an official version of the schema.org schemas at http://schema.org/docs/schemaorg.owl

“This allows the schema.org schemas to be used with all OWL-aware tools such as editors, validators etc., as well as to create mappings to other Semantic Web schemas.

We would like to acknowledge the Linked Data Research Center at DERI, in particular Michael Hausenblas and Richard Cyganiak, for their work on schemas.rdfs.org, and for their help in developing the OWL schema for schema.org.”
A lot of Events Categories in Schema.org

THE WORLD OF SCHEMA.ORG

// Power-user tips // 1. Hold “Alt” to expand/collapse recursively a node (try it on the root)! // 2. Click on the right column to go to the relevant schema.org documentation //</p>
take away message
don't bury
your data in some HTML page
when you publish a page that contains data...
do make the embedding explicit
Linked Data Principles

- Tim Berners Lee [2006] (*Design Issues*)

1. Use URIs to identify things (anything, not just documents);
2. Use HTTP URIs – globally unique names, distributed ownership – so that people can look up those names;
3. Provide useful information in RDF – when someone looks up a URI;
4. Include RDF links to other URIs – to enable discovery of related information
An Example: DBpedia

- DBpedia is a community effort to:
  - extract structured "infobox" information from Wikipedia
  - interlink DBpedia with other datasets on the Web
Scraping infobox data

About: Bogotá
An Entity in Data Space: dbpedia.org

Bogotá (anciennement Santa fé de Bogotá Distrito Capital) est la capitale de la Colombie et du département de Cundinamarca.

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbpedia-owl.areaLand</td>
<td>668.7</td>
</tr>
<tr>
<td>dbpedia-owl.areaTotal</td>
<td>1587</td>
</tr>
<tr>
<td>dbpedia-owl.elevation</td>
<td>2640</td>
</tr>
<tr>
<td>dbpedia-owl.establishedTitle</td>
<td>Foundation</td>
</tr>
<tr>
<td>dbpedia-owl.foundingDate</td>
<td>1538-08-06 (xsd:date)</td>
</tr>
<tr>
<td>dbpedia-owl.foundingPerson</td>
<td>dbpedia:Gonzalo_Jim%C3%A9nez_de_Quesada</td>
</tr>
<tr>
<td>dbpedia-owl.leaderName</td>
<td>dbpedia:Alternative_Democratic_Pole</td>
</tr>
<tr>
<td></td>
<td>dbpedia:Samuel_Moreno_Rojas</td>
</tr>
<tr>
<td>dbpedia-owl.leaderTitle</td>
<td>Mayor</td>
</tr>
<tr>
<td>dbpedia-owl.motto</td>
<td>Bogotá, 2600 meters closer to the stars</td>
</tr>
<tr>
<td></td>
<td>Bogotá, 2600 metros más cerca de las estrellas</td>
</tr>
<tr>
<td>dbpedia-owl.populationAsOf</td>
<td>2005</td>
</tr>
<tr>
<td>dbpedia-owl.populationDensity</td>
<td>22593</td>
</tr>
<tr>
<td></td>
<td>11071</td>
</tr>
<tr>
<td>dbpedia-owl.populationMetro</td>
<td>7881156</td>
</tr>
<tr>
<td>dbpedia-owl.populationTotal</td>
<td>6776009</td>
</tr>
</tbody>
</table>

http://dbpedia.org/resource/Bogotá
Automatic Links Among Open Datasets

<http://dbpedia.org/resource/Bogotá>
  owl:sameAs <http://sws.geonames.org/3688689/>
  owl:sameAs
  <http://rdf.freebase.com/ns/guid.9202a8c04000641f8000000000167bab>
  dbpedia:population "6776009"
...

<http://sws.geonames.org/3688689/>
  owl:sameAs <http://dbpedia.org/resource/Bogotá>
  wgs84_pos:lat "4.6"
  wgs84_pos:long "-74.0833333"
  geo:population "7102602"
...

DBpedia

Geonames
The Web of Data has many equivalent URIs. This service helps you to find co-references between different data sets. Enter a known URI, or use Sindice to search first.

Equivalent URIs for http://dbpedia.org/resource/Bogotá –

5. http://telegraphis.net/data/capitals/CO/Bogotá#Bogotá

rdf+xml · n3 · json · text · show fewer items
Bogotá on Freebase

Bogotá

Bogotá (Spanish: Bogotá [ˈbogota]; officially named Bogotá, D.C. (D.C. for "Distrito Capital", which means "Capital District"), formerly called Santa Fe de Bogotá – is the capital city of Colombia, as well as the most populous city in the country, with 6,776,009 inhabitants (2005). Bogotá and its metropolitan area, which includes municipalities such as Chía, Cota, Soacha, Cajicá and La Calera, had an estimated population of 8,244,980. In te...

Read article at Wikipedia

Country: Colombia

Area: 1,587 km²

Also known as: Bogota, Bogotá, Colombia, Distrito Capital De Bogotá, Distrito Capital De Bogota, Distrito Capital De Santafé De Bogotá, Bogota Distrito Especial, Bogotá Distrito Especial, Bogotade, Distrito Capital De Santafe De Bogota, Special District Of Bogota, Bogota D.E., Special District Of Bogotá, Distrito Especial De Bogotá, Distrito Capital, Distrito Especial, Bogotá D.E., Distrito Especial De Bogota, Special Capital District
Bogotá on Geonames

Map center: N 4° 36' 25" W 74° 4' 59"

Bogotá ca. 2620 m

Bogotá, Bogota, Bogotá, Santa Fe de Bogotá...
Colombia » Bogotá D.C.
capital of a political entity
population: 7102602
N 4° 36' 0" W 74° 5' 0"
4.6 / -74.08333
GeoNameId: 3688689

refresh to display all features in area

GeoNames Wikipedia
How Much Linked Data is there?

As of May 2007
Linked Data Cloud – August 2007
Linked Data Cloud – March 2008
Linked Data Cloud – September 2008

As of September 2008
The Web of Data

- Expose open datasets in RDF
- Set RDF links among the data items for different datasets
- Over 26 billion triples, 500 millions links, 203 datasets (September 2010)
- ... still counting
… but let’s STOP counting!

- Linked Open Numbers
  (April 1\textsuperscript{st} 2010)

- Linked Open Colors
  (April 1\textsuperscript{st} 2011)
  \url{http://purl.org/colors/}

Linked Open Numbers

Some examples:
- \url{http://purl.org/colors/rgb#FF0000}
- \url{http://purl.org/colors/rgb#00FF00}
- \url{http://purl.org/colors/rgb#0000FF}
- \url{http://purl.org/colors/rgb#FF00FF}
- \url{http://purl.org/colors/rgb#00FF00}

The source code is available at \url{http://www.improvise.net/loc}.
Linked data summary

- URIs, possibly identifying media fragments
- + annotations (tags)
- + links among fragments & annotations

geonames: 2950159

nar: location

nar: subject

nc: 15054000

foaf: depicts

dbpedia: Zidane

wp: 2006_FIFA_World_Cup#Final

events: id
Searching Entities in the Cloud

Entity Search, Find, and Explore

Faceted Search & Find Service

(c) 2008, Georgi Kolev, Freie Universität Berlin
The Beatles (Band)

The Beatles were an English rock and pop group formed in Liverpool in 1960 who became one of the most commercially successful and critically acclaimed bands in the history of popular music. During their years of stardom, the group consisted of John Lennon (rhythm guitar, vocals), Paul McCartney (bass guitar, vocals), George Harrison (lead guitar, vocals) and Ringo Starr (drums, vocals).
Searching Linked Data

Swoogle
semantic web search 2007

indice
THE SEMANTIC WEB INDEX

watsön
exploring the semantic web

Falcons
Sindice already crawling Schema.org
WebKit and Chrome prerendering

Google search result pages now trigger a prefetch of top search result links in an effort to make navigating search results as easy as changing channels on your television. If Google's search algorithms determine there is a significant probability of user click-through on particular result they will instruct supporting browsers to preload the entire destination page including images, JavaScript, advertisements, and analytics. Update your web pages to be aware of the current page visibility state and track interactions, not background tasks.

Read more...

SSL statistics from Chrome and Googlebot

The Google Chrome team released new statistics and implementation details on their proposed "False Start" abbreviated TLS handshake. Google claims the new handshake, introduced in version 9 of the Chrome browser in February, shaves an average of 120 milliseconds from a typical four-flight TLS handshake by accepting application data before both sides have communicated a "Finished" status.

Read more...
Browsing Linked Data

Example starting places are below. You can add new ones by putting the URI in the address bar above and clicking "Add to outliner".

Key: Data: ○ fetch data, ● loading, ● failed, ● loaded. Display: ▶ expand, ▼ collapse,
Browsing Linked Data

**Disco**
*(Free University of Berlin)*

- Label of the displayed resource
- Navigation box
- Resource description
- List of all source graphs
- Link for displaying the session cache
- Go to URI button
- Sources of each piece of information
Browsing Linked Data

Marbles
(Free University of Berlin)
Browsing Linked Data

Zitgist
(Zitgist LLC)
Browsing Linked Data

OpenLink Data Explorer
(OpenLink Software)
VisiNav

**Angela Merkel**

http://dbpedia.org/resource/Angela_Merkel

person, Spatial Thing, organization

is the Chancellor of Germany. Merkel, elected to the German Parliament from Mecklenburg-Vorpommern, has been the chairwoman of the Christian Democratic Union since 9 April 2000, and Chairwoman of the CDU-CSU parliamentary party group from 2002 to 2005. ...

**Angela Merkel on Flickr - Photo Sharing!**

http://www.flickr.com/photos/77796851@N00/41699553

Document Resource Document

Angela Merkel political poster with an "Ich fühle mich so leer." sticker. The sticker was taken from a garbage can ad campaign.

**Angela Merkel auf der CeBIT, 2002 on Flickr - Photo Sharing!**

http://www.flickr.com/photos/79925326@N00/10849705

Document Resource Document

Angela Merkel auf der CeBIT, 2002
freebase parallax

a novel way to browse and explore data

August 2008

David Huynh, Research Scientist, david@metaweb.com
Metaweb Technologies, Inc., http://metaweb.com
TimBL Vision back in 1994
Web pages describe the World

Each makes ‘claims’

They can disagree
FOAF is a project about sharing information in the Web. It's about ways of describing things using computers, so that those descriptions can be linked together, mixed up with other data, and searched.

Friend of a Friend

Henry says, “My name is ‘Henry Story’”

Joe says, “I know Henry who knows Jane”

Joe knows someone called “Henry Story”
FOAF (Friend-of-a-Friend)

- FOAF is an ontology for describing people and the relationships that exist between them
- Can be integrated with any other SW vocabularies
- Some services with FOAF exports:
  - hi5
  - LiveJournal
  - VOX
  - tribe
- People can also create their own FOAF document and link to it from their homepage
- FOAF documents usually contain personal info, links to friends, and other related resources
The FOAF Specification

Integrating SN with FOAF for reuse

* Source: Sheila Kinsella, Applications of Social Network Analysis 2007
Going through the Walled Gardens

FOAF Naut

foafnaut

[Image of a FOAF network diagram]
Welcome to the FOAF builder

FOAF (Friend of a Friend) is a way of describing people, their activities, and their relationships to other people and things. FOAF allows groups of people to truly open social networks that are controlled by individual users and not by any one single entity.

Garlik’s FOAF builder is a user-friendly way of allowing you to create your own or edit an existing FOAF profile.

To create your own simply click 'build', or you can add details to your FOAF profile automatically by providing usernames of your Flickr, Last.fm or LiveJournal accounts. If you already have a FOAF profile you can enter it in below.

Flickr username:  

Last.fm username:  

Livejournal username:  

FOAF profile:  

Build

If you have saved your FOAF file with us, please click logon
XML Watch: Finding friends with XML and RDF

The Friend-of-a-Friend vocabulary can make it easier to manage online communities

Edd Dumbill (edd@xml.com), Editor and publisher, xmlhack.com

Summary: Edd Dumbill explores an XML and RDF application known as Friend-of-a-Friend (FOAF). FOAF allows the expression of personal information and relationships, and is a useful building block for creating information systems that support online communities. Code samples demonstrate the basics.

View more content in this series
Relationship Vocabulary

- **http://purl.org/vocab/relationship** (Apr 2010)

  acquaintance of  child of  collaborates with  lost contact with
  ambivalent of  apprentice to  close friend of  colleague of  employed by
  ancestor of  enemy of  has met  influenced by  knows by reputation
  grandchild of  knows in passing  lives with  mentor of  neighbor of  parent of
  participant  relationship  sibling of  spouse of  works with  would like to now

- 35 new properties to complement FOAF
Semantically-Interlinked Online Communities

- A schema for representing users, forums, posts and threads, containers, and other items in online community sites, for reuse and interoperability:
  - Aims to fully describe the structure of content in these sites
  - Also to create new connections between forums and posts from different types of discussion systems (blogs, forums, mailing lists, etc.) and content items/containers on Web 2.0 sites
  - And to browse connected posts and channels in interesting ways (e.g. distributed linked conversations, decentralised discussion channels and communities, etc.)
The SIOC ontology

Producing SIOC data

- Over 20 applications for producing SIOC data:
  - Many are free and open source
  - Blogs and forums: WordPress, phpBB, Drupal, b2evolution
  - "Legacy" applications: mailing lists, IRC
  - New media: Twitter, Jaiku, Facebook, Flickr

- APIs for those who may wish to make their own producers:
  - PHP, Perl, Java, Ruby on Rails
Portable Data with SIOC and FOAF

People
- Alice
- Bob
- Carol

Services and User Accounts
- foaf:Person
- foaf:knows
- foaf:knows
- foaf:knows
- foaf:knows

Containers and Content Items
- Roberto’s Bookmarks
- DataPortability.org
- WebCamp.org
- Robert Rodriguez’s Blog
- Happy New Year!
- My Christmas Wish List...

EURECOM
Collect SIOC from various sources

Looking at 42 sioc:posts

Is AtomPP for or not for blog?

Rob Yates is bringing up a serious point on the work of the Atompub WG:

The Atom Publishing Protocols flexibility is its real strength, but it seems like an omission if the working group declares ...

Cisco’s social network acquisitions (Was Sunday Tribune: “Social networking: sharing is caring for the bottom line”)

posted by Cloud 1 day ago (March 20, 2007 11:08) — go to original post

I was quoted for this article in the Sunday Tribune by Damien about our boards.ie CI (commercial interaction) forums. There are other interesting parts to the article, in particular Cisco’s move into the social networking domain earlier this month. When I wrote earlier about Cisco’s “Human Network” ad, I hadn’t realised that Cisco had actually purchased Tribe.net (and Five Across), thinking that they were content to simply have us use their networking infrastructure for whatever collaborative purposes we enjoyed. But despite much bewilderment, there seems to be some good reasons for this acquisition (see Marc Canter’s list and Winer’s tongue-in-cheek take on the move). (From the Semantic Web side, maybe Cisco could be persuaded to provide further convergence via OpenID, Tribe’s FOAF provisions and something similar from Five Across’ “Connect” product.)
Consuming SIOC via Exhibit
Dublin Core

- **http://purl.org/dc/elements/1.1/** (Jan 2008)

- **15 elements or attribute-value pairs (simple DC)**
  - Contributor, Coverage, Creator, Date, Description, Format, Identifier, Language, Publisher, Relation, Rights, Source, Subject, Title, Type

- **55 elements or attribute-value pairs (qualified DC)**
  - [http://purl.org/dc/terms/](http://purl.org/dc/terms/)
Dublin Core example

<dc:title>Washing & ironing clothes.</dc:title>
<dc:date>ca. 1942</dc:date>
<dc:description>Mexican workers washing and ironing clothes.</dc:description>
<dc:subject>Agricultural laborers--Mexican--Oregon; Agricultural laborers--Housing--Oregon; Laundry</dc:subject>
<dc:type>Image</dc:type>
<dc:source>Silver gelatin prints</dc:source>
<dc:rights>Permission to use must be obtained from OSU Archives.</dc:rights>
<dc:identifier>P20:1069</dc:identifier>
<dc:identifier>http://digitalcollections.library.oregonstate.edu/u?/bracero,37</dc:identifier>
Good Relations

- gr:BusinessEntity for a company or business,
- gr:LocationOfSalesOrServiceProvisioning for a store,
- gr:ProductOrServicesSomeInstancesPlaceholder for products or services (if there are multiple items),
- gr:ActualProductOrServiceInstance for a particular product or service (e.g. used items),
- gr:ProductOrServiceModel for the datasheet describing the features of a product, and
- gr:Offering for an offer to sell, repair, lease something, or to express interest in such an offer.
The GoodRelations Ontology for E-Commerce
Language Overview - UML Class Diagram

http://purl.org/goodrelations/

Version 1, Release 2010-04-12
Martin Hepp, mhepp@computer.org

Notes:
1. The following GoodRelations elements are not shown in this diagram because they are only shortcuts for simpler annotation or querying. See the documentation at http://purl.org/goodrelations/ for details:
   - gr:hasMaxValue (shortcut for setting both hasMinValue=false and hasMaxValue=true properties to the same value in one turn)
   - gr:hasMaxValue (shortcut for setting both hasMinValue=false and hasMaxValue=false properties to the same value in one turn)
   - gr:hasMinValue (shortcut for setting both hasMinValue=false and hasMaxValue=false properties to the same value in one turn)

2. Also, the class gr:MyRelations is not shown, because it is just a helper class to collate all classes that represent many relations that OWL cannot handle otherwise.

For the recommended cardinality of attributes, see the GoodRelations Language Reference at http://purl.org/goodrelations/v1.html.

Red highlighting indicates elements added or changed in this release.
At last year's SemTech conference, Myers said that it had resulted in a 30% increase in search traffic.
The Open Graph Protocol
Open Graph: Getting Started

<html xmlns:og="http://opengraphprotocol.org/schema/">
  xmlns:fb="http://www.facebook.com/2008/fbml">
  <head>
    <title>The Rock (1996)</title>
    <meta property="og:title" content="The Rock"/>
    <meta property="og:type" content="movie"/>
    <meta property="og:url" content="http://www.imdb.com/title/tt0117500/"/>
    <meta property="og:image" content="http://ia.media-imdb.com/rock.jpg"/>
    ...
  </head>
  ...
</html>
Open Graph Properties

- The Open Graph protocol defines 5 required properties:
  - og:title - The title of your object as it should appear within the graph, e.g., "The Rock".
  - og:type - The type of your object, e.g., "movie". See also http://developers.facebook.com/docs/opengraph#types
  - og:image - An image URL which should represent your object within the graph. The image must be at least 50px by 50px and have a maximum aspect ratio of 3:1.
  - og:url - The canonical URL of your object that will be used as its permanent ID in the graph, e.g., http://www.imdb.com/title/tt0117500/
  - og:site_name - A human-readable name for your site, e.g., "IMDb"

- Optional properties
  - og:description - A one to two sentence description of your page.*
  - + location (7 properties) + contact (3 properties)
rNews for the Press

- RDFa vocabulary for news articles
  - Easier to implement than NewsML
  - Easier to consume for news search and other readers, aggregators

- Under development at the IPTC
  - March: v0.1 approved
  - Final version by Sept
Wrap up: popular vocabularies
Agenda

- A crash course in the world of structured data
  - #microdata, #microformat, #rdfa
  - #rdf, #owl, #skos, #sparql, #linkeddata

- EventMedia (User-centered design approach)
  - LODE: a model for representing events
  - Scraping and interlinking description of events
  - Enriching events with illustrating media
  - Detecting events from social media activities

- Detecting events from human sensing
  - #twitter, #foursquare, #facebook
What are Events?

Events are observable occurrences grouping People, Places, and Time.
Ontology: Making an abstraction

What? Where? When? Who?
Event-based centric interfaces

- Action or occurrence taking place at a certain time at a specific location
  - Useful for organizing and browsing collections of media
  - Useful for discovering complex relationships between data

⇒ Need for an expressive event model for connecting pieces of data

⇒ Not Yet Another Model!
There are already many event ontologies

<table>
<thead>
<tr>
<th>Event Model</th>
<th>Ontology URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIDOC CRM</td>
<td><a href="http://cidoc.ics.forth.gr/OWL/cidoc_v4.2.owl">http://cidoc.ics.forth.gr/OWL/cidoc_v4.2.owl</a></td>
</tr>
<tr>
<td>ABC Ontology</td>
<td><a href="http://metadata.net/harmony/ABC/ABC.owl">http://metadata.net/harmony/ABC/ABC.owl</a></td>
</tr>
<tr>
<td>Dolce+DnS Ultralite</td>
<td><a href="http://www.loa-cnr.it/ontologies/DUL.owl">http://www.loa-cnr.it/ontologies/DUL.owl</a></td>
</tr>
<tr>
<td>F</td>
<td><a href="http://events.semantic-multimedia.org/ontology/2008/12/15/model.owl">http://events.semantic-multimedia.org/ontology/2008/12/15/model.owl</a></td>
</tr>
</tbody>
</table>
Fundamental Types of Events

- **Aspect**: ongoing activity vs transition between states
  - \( \text{cyc: Event} \cap \text{cyc: StaticSituation} \leq \text{cyc: Situation} \)
  - \( \text{cidoc:E5.Event} \cap \text{cidoc:E3.Condition_State} \leq \text{cidco:E2.Temporal_Entity} \)
  - \( \text{abc:Event is a transition between abc:Situation} \approx \text{cidoc:E3.Condition_State} \)

- **Agentivity**: who has produced the event?
  - \( \text{cyc:Action, dul:Action} \leq \text{Event} \)
  - \( \text{E7.Activity} \leq \text{E5.Event} \)
  - \( \text{abc:Action} \cap \text{abc:Event} = \emptyset \)
    - Events are fully described as a set of actions taken by specific agents
    - Issue for modeling e.g. earthquakes

- **Interpretation matters!**
  - Identifiable changes or not? Agency can be assigned?
  - \( \text{dul:Situation describe dul:Event} \)
  - \( \text{dul:Action, dul:Process} \leq \text{dul:Event} \)
Events and Temporal Intervals

- Relating events to chronological spans of time
  - Persistent, socially attributed meanings
  - Arbitrary system for subdividing an abstract space

- Modeling a class for temporal intervals and use an OP
  - ABC, CIDOC, EO (owl:TemporalEntity)

- Modeling a XML Schema typed value and use a DP
  - Pro: simplicity, values expressed as xsd:date or xsd:dateTime
  - Cons: inability to express uncertain period or when there is no coincidence with date units

- Having two properties
  - dul:hasEventDate ... literal value
  - dul:isObservableAt ... dul:TimeInterval
Events, Spaces and Places

- **Relating events to places**
  - Semantically significant places
  - Abstract spatial regions

- **Support spatial regions only: ABC, CIDOC, EO**
  - eo: Event → eo: place → wgs84: SpatialThing

- **Support the place/space distinction**
  - dul: Event → dul: hasLocation → dul: Place
  - dul: Event → dul: hasRegion → dul: SpaceRegion
  - Most flexible approach: allow to resolve to places with no geographical coordinate systems (e.g. mythical events, SecondLife)
Participation in events

- **Object involvement in events:**
  - Simple involvement in event:
    - $\text{abc:Event} \rightarrow \text{abc:involves} \rightarrow \text{owl:Thing}$ ($\leq \text{abc:Actuality}$)
    - $\text{cidoc:E5.Event} \rightarrow \text{cidoc:P12.occurred_in_the_presence_of} \rightarrow \text{cidoc:E77}$
    - $\text{dul:Event} \rightarrow \text{dul:hasParticipant} \rightarrow \text{dul:Object}$
    - $\text{eo:Event} \rightarrow \text{eo:factor} \rightarrow \text{owl:Thing}$
  - Tangible thing which results from an event:
    - $\text{abc:Event} \rightarrow \text{abc:hasResult} \rightarrow \text{owl:Thing}$
    - $\text{eo:Event} \rightarrow \text{eo:product} \rightarrow \text{owl:Thing}$

- **Agent participation in events:**
  - $\text{abc:hasParticipant} \leq \text{abc:hasPresence}$
  - $\text{cidoc:P11.had_participant} \leq \text{cidoc:P14.carried_out_by}$
  - $\text{dul:involvesAgent} \leq \text{abc:hasParticipant}$
Events, Influence, Purpose and Causality

- Making broad assertions linking events to any thing
  - cidoc:P12.occurred_in_the_presence_of, cidoc:P15.was_influenced_by
  - eo:factor, abc:hasResult

- F model uses the DnS pattern
## Events, Parts and Composition

- **Event A being part of event B ≠ A's timespan ∈ B's timespan**
  - `cidoc:P86.falls_within` for expressing containment among timespans
  - `cidoc:P9.consist_of` ≈ `eo:sub_event` ≈ `abc:isSubEventOf`

- **Linking sub-events with parthood**
  - `dul:hasPart`
    - *The 20th century contains the year 1923*
    - *World War II included Pearl Harbour*

- **Linking sub-events with composition**
  - `dul:hasConstituent`
    - *The French revolution is composed of the Bastille catch*
Towards a Linked Data Event Model
# Some mappings in LODE

<table>
<thead>
<tr>
<th>ABC</th>
<th>CIDOC</th>
<th>DUL</th>
<th>EO</th>
<th>LODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>atTime</td>
<td>P4.has_time_span</td>
<td>isObservableAt</td>
<td>time</td>
<td>atTime</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P7.took_place_at</td>
<td></td>
<td>place</td>
<td>inSpace</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inPlace</td>
<td></td>
<td>hasLocation</td>
<td></td>
<td>atPlace</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>involves</td>
<td>P12.occurred_in_the_presence_of</td>
<td>hasParticipant</td>
<td>factor</td>
<td>involved</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hasPresence</td>
<td>P11.had_participant</td>
<td>involvesAgent</td>
<td>agent</td>
<td>involvedAgent</td>
</tr>
</tbody>
</table>
LODE: An ontology for Linking Open Descriptions of Events

This Version
http://linkedevents.org/ontology/2009-07-28/ [HTML] [RDF/XML]

Latest Version
http://linkedevents.org/ontology/

Authors
Ryan Shaw

Contributors
Raphaël Troncy
Lynda Hardman

Copyright © 2009 Ryan Shaw

This work is licensed under a Creative Commons License.

Table of Contents

- Introduction
- Namespace
- Summary of Terms
- Vocabulary Classes
- Vocabulary Properties
- License

Introduction
Representing Events with

 LOD: Linking Open Descriptions of Events

 lode:Event
  lode:concert
  lode:atPlace: dbpedia: The_Henry_Fonda_Theater
  lode:atTime: lode:illustrate
  lode:inSpace: lode:involvedAgent
  geo:Point
    geo:lat: 34.101944°N
    geo:long: -118.340972°W
  lode:time: 2010-01-24T20:00:00
    xsd:date:time
    http://www.youtube.com/watch?v=yHHZGQRDKJA
  lode:creator: http://www.youtube.com/user/aghorrorag

29/06/2011 - Lecture at the 2nd Summer School on Social Media Retrieval (S3MR) - Antalya, Turkey
EventMedia Goals (User-Centered Design)

1. Discover PAST, PRESENT and FUTURE events
2. Live, relive and predict experiences through shared media
3. Identify meaningful and/or interesting relationships between events/media/people
1st Collect some opinions...

Online Survey (n=28), 2 group discussions (n=35)

Past Experiences (Memorable Events)
- Discovery
- Decision making
- Registering & sharing
- Meaningful relationships

Existing Technologies
- Opinions
- Interests
- Suggestions
- Benefits/drawbacks

Scenarios Requirements
1st Design Concept
EventMedia Project: Questionnaire

1. Think about a memorable/recent event you have participated:
   ➢ Tell us what it was and what type of event was it

2. How do you usually find out or look for such events?

3. What is important to support your decision about going to an event?

4. Once you attended to an event, how do you register the moment and share your experience?

5. What could be considered meaningful (surprising or entertaining) relationships among events?
Brainstorm online with users

Brainstorm in Wave (EventMedia Project)

So I thought I would shake things up and see if this Brainstorm in wave works. Why? Because its asynchronous, it handles well discussions and you can keep adding information as you remember.

What is it about? I'm currently leading an explorative study within Petmedia Network of Excellence to identify Interface and Interaction aspects of an event-based environment for Media exploration. More specifically, the project focuses on using social peer 2 peer, media content analysis and semantic web to identify meaningful relationships among events, and represent them through videos and pictures. In this phase of the study I'm exploring potential concepts and trying to identify what are the variables and user interests that have to be considered when dealing with events so that they can be further explored in following iterations.

How will it work?

Duration: The Brainstorm will start tomorrow (Thursday 14th) morning and end on Friday (15th) afternoon. This way you can fill in the answers on your own time.

Will be up for a few more days to give time for more people to participate.

Questions: I will leave some questions throughout the day. This will be divided into 2 sessions, the first concerning past experience (events) and the second to explore interests and possibilities (services). If the question is unclear, please ask. Keep in mind that questions are subjective and you may answer whatever you wish (there are no right answers).

Answers: Try to keep your answers short and straightforward (pretend it's as long as an SMS or Twitter message). Also, if you think somebody replied something...
2nd Look into “real” behaviors...

- Scenario based study (2 sessions, n=15)

Scenarios

Reenact

Opinions

29/06/2011 - Lecture at the 2nd Summer School on Social Media Retrieval (S3MR) - Antalya, Turkey
Behavioral Patterns

- **Discovery**
  - Invitations and recommendations
  - Rely on traditional media
  - Social networks (facebook - students)
  - Previously attended events or venues

- **Decision Making**
  - Who’s Joining?
  - Where, When, How Much? (constraints)
  - What? (e.g. type, performer, topic)
  - Subjective factors (fun, atmosphere)
**Behavioral Patterns**

- **Registering and Sharing**
  - Communicating their experience
  - Pictures and short videos (for sharing)
  - Media directories and social networks

- **Meaningful Relationships**
  - Similar categories, attributes and content
  - User attendance (similar interests, behaviors)
  - Repeated events (e.g. annual festivals)
Behavioral Patterns

![Diagram showing relationships between Google, eventful, Facebook, YouTube, flickr, and Twitter]

29/06/2011 - Lecture at the 2nd Summer School on Social Media Retrieval (S3MR) - Antalya, Turkey - 172
Existing Services

- Single source with overview (?)
- Allows opportunistic/serendipitous discovery
- Limited exploration/browsing features
- Information overload (cluttered, difficult)
- Information incompleteness (coverage, decision)
Organize the mess

- Scrape event directories
- Link the information
- Find media illustrating events
- Design the application Interface

Representing Events with LODE

Linking Open Descriptions of Events

- iode:Event
- time:Instant
  - time:inXSDDateTime
    - 2007-11-13T20:30:00
  - ^xsd:dateTime
- iode:Concert_Live_Music
  - rdf:type
  - iode:atTime
  - iode:illustrate
  - http://www.youtube.com/watch?v=pOAUmwXfDZI
- http://foursquare.com/venue/185188
  - iode:atPlace
  - iode:inSpace
  - owl:sameAs
  - geo:Point
    - 48.865865^N
    - 2.3778^W
- dbpedia:Róisín_Murphy
  - ma:creator
  - http://www.youtube.com/user/cartoixa
Linking the Data
Reasoning & Annotation

- Time, Location and Attendance

![flickr](image1.png)
![eventful](image2.png)
![foursquare](image3.png)
Collaborative Filtering

- Disambiguate and propagate information about attendance
- Identify Interests and provide Recommendations

Customers Who Bought This Item Also Bought

- **Semantic Web for the Working Ontologist: Effect...** by Dean Allemang
  - Rating: ★★★★☆ (2)
  - Price: £30.39

- **Semantic Web Programming** by Mike Dean
  - Price: £20.14

- **Semantic Web For Dummies** by Jeffrey T. Pollock
  - Rating: ★★★★☆ (1)
  - Price: £16.99

- **Programming the Semantic Web** by Toby Segaran
  - Rating: ★★☆☆☆ (1)
  - Price: £15.53
Interlinking

- Linking Agents with
  - Freebase, Dbpedia, MusicBrainz

- Linking Venues with
  - Geonames, Dbpedia, Foursquare (via Uberblic)

- Linking Events with
  - Last.fm, Upcoming, Eventful

- Linking Categories with
  - Facebook, Eventful, Upcoming, Zevents, LinkedIn, Eventbrite, TicketMaster

- Linking Users with
  - Social Graph API
Interlinking

**Event**
- Last.fm
- Eventful
- Upcoming
- DBpedia
- Freebase
- Uberblic

**Location**
- Last.fm
- Eventful
- Upcoming
- DBpedia
- Freebase
- Foursquare
- Geonames

**Agent**
- Last.fm
- Eventful
- MusicBrainz
- DBpedia
- Freebase
- Uberblic
- New York Times
**SILK Framework**

- Based on the Silk-LSL link specification language
- Transformation and algebraic functions: max, min, avg, etc.
- Several metrics available:
  - Syntax: equality, Jaro, Leveinstein, n-gram
  - Lexical: WordNet
  - Geo: wgs84
  - Temporal: date
Silk Framework

Configuration

SILK - LSL

```xml
<LinkCondition>
  <Aggregate type="average">
    <Compare metric="jaro">
      <TransformInput function="lowerCase">
        <Input path="?a/rdfs:label" />
      </TransformInput>
      <TransformInput function="lowerCase">
        <Input path="?b/rdfs:label" />
      </TransformInput>
    </Compare>
    <Compare metric="wgs84">
      <TransformInput function="concat">
        <Input path="?a\lode:atPlace/lode:inSpace/wgs84:lat" />
        <Input path="?a\lode:atPlace/lode:inSpace/wgs84:long" />
        <Param name="glue" value=" " />
      </TransformInput>
      <TransformInput function="concat">
        <Input path="?b\lode:atPlace/lode:inSpace/wgs84:lat" />
        <Input path="?b\lode:atPlace/lode:inSpace/wgs84:long" />
        <Param name="glue" value=" " />
      </TransformInput>
      <Param name="unit" value="km"/>
      <Param name="threshold" value="10" />
    </Compare>
  </Aggregate>
</LinkCondition>
<Filter threshold="0.8" />
```
Alignement for Agents

- Alignement base on:
  - foaf:Agent
  - rdfs:label

<table>
<thead>
<tr>
<th></th>
<th>Eventful (6543)</th>
<th>Last.fm (50151)</th>
<th>MusicBrainz (459023)</th>
<th>Dbpedia (107112)</th>
<th>Uberblic (236691)</th>
<th>NYTimes (4794)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eventful</td>
<td>-</td>
<td>2865 (44%)</td>
<td>3616 (55%)</td>
<td>1985 (30%)</td>
<td>1567 (24%)</td>
<td>7 (0.1%)</td>
</tr>
<tr>
<td>Last.fm</td>
<td>2865 (6%)</td>
<td>-</td>
<td>26619 (53%)</td>
<td>9442 (19%)</td>
<td>12905 (26%)</td>
<td>14 (0.03%)</td>
</tr>
</tbody>
</table>

- Examples:
  - Donavan Frankenreiter / Donovan Frankenreiter (Jaro 0.98)
  - Birds & Batteries / Birds and Batteries (Jaro 0.70)

- Total:
  - Eventful : 61 %
  - Last.fm : 58 %
Alignement for Locations

<table>
<thead>
<tr>
<th></th>
<th>Eventful (13516)</th>
<th>Last.fm (15857)</th>
<th>Upcoming (5173)</th>
<th>DBpedia (496728)</th>
<th>Foursquare (641770)</th>
<th>Geonames (1090357)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eventful</td>
<td>-</td>
<td>998 (7%)</td>
<td>366 (3%)</td>
<td>90 (0.7%)</td>
<td>1296 (10%)</td>
<td>320 (2%)</td>
</tr>
<tr>
<td>Last.fm</td>
<td>998 (6%)</td>
<td>-</td>
<td>626 (4%)</td>
<td>141 (0.9%)</td>
<td>911 (6%)</td>
<td>345 (2%)</td>
</tr>
<tr>
<td>Upcoming</td>
<td>366 (7%)</td>
<td>626 (12%)</td>
<td>-</td>
<td>74 (1.4%)</td>
<td>1300 (25%)</td>
<td>232 (4%)</td>
</tr>
</tbody>
</table>

- **Examples:**
  - ✔️ The Stone **Bar** (34.1019 ; -118.304) \( Dist : 29 \text{ m} \) – \( \text{Score (sim)} : 0.98 \)
  - ✔️ The Stone (34.1017 ; -118.304)
  - ✗ fall harvest wine dinner **bavarian inn restaurant frankenmuth** (43.32 ; -83.73) \( Dist : 80 \text{ m} \) \( \text{Score} : 0.92 \)
  - ✗ **Frankenmuth Bavarian Inn Restaurant**

- **Total:**
  - ✔️ Eventful : 17 %
  - ✔️ Last.fm : 15 %
  - ✔️ Upcoming : 36 %
Alignement for Events

- Alignement based on title, location and time

<table>
<thead>
<tr>
<th></th>
<th>Eventful</th>
<th>Last.fm</th>
<th>Upcoming</th>
<th>DBpedia Music Festival</th>
<th>Uberblic Performer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eventful</td>
<td>-</td>
<td>76 (0,2%)</td>
<td>34 (0,1%)</td>
<td>28 (0,1%)</td>
<td>15 (0,04%)</td>
</tr>
<tr>
<td>Last.fm</td>
<td>76 (0,1%)</td>
<td>-</td>
<td>586 (1%)</td>
<td>389 (0,7%)</td>
<td>1148 (2%)</td>
</tr>
<tr>
<td>Upcoming</td>
<td>34 (0,3%)</td>
<td>586 (4%)</td>
<td>-</td>
<td>31 (0,2%)</td>
<td>15 (0,1%)</td>
</tr>
</tbody>
</table>

- Example:
  - LastFm : "Camp Bestival" à "Lulworth Castle" le 18/07/2008
  - Eventful : "New Camp Bestival Dorset" à "Lulworth Castle" le 18/07/2008

- Total:
  - Eventful : 0,4 %
  - Last.fm : 3,8 %
  - Upcoming : 4,8 %
Research challenges

http://oaei.ontologymatching.org/2011/
What are Events?

Events are observable occurrences grouping

People  Places  Time

Experiences documented by Media
Radiohead / Thom Yorke _RH03079xLAR

Radiohead for HAITI BENEFIT @ the Henry Fonda Theater.
Set list & show info on the previous photo: www.flickr.com/photos/hazyskyline/4303435326

I wish I could formulate my thoughts better & write some sort of meaningful note here, but really every time I try its just filled with words like "love, amazing, unbelievable, epic, holy shit, wow, perfect, fantastic, etc."
so for right now, I'm just sticking to tags instead!

hazyskyline.wordpress.com/

Do not steal or repost my photos without permission and proper credit. To purchase prints, contact me.

Commentaires et favoris

grinypoodes (1 y ± 7 mois)
Um, wow. Look at those awesome colors!

kirstiecat (1 y ± 7 mois)
I love this mood... he almost looks like a lost soul who is about to fall right into his mic
Media explicitly associated with the event

29/06/2011 - Lecture at the 2nd Summer School on Social Media Retrieval (S3MR) - Antalya, Turkey
Representing Media with Media Ontology

http://farm3.static.flickr.com/2038/2014900103_d2b12a3c29_z.jpg

2007-11-13T21:18
^^xsd:dateTime

Róisín Murphy - Pandora

http://www.youtube.com/watch?v=pOAUmXfDZ

http://www.flickr.com/people/cartoxa/

http://www.youtube.com/user/cartoxa

concert

ma:image

flickr:2014900103

240^^xsd:integer

160^^xsd:integer

ma:keyword

foaf:thumbnail

ma:locator

ma:CreateDate

lode:illustrate

sioc:hasCreator

ma:Video

ma:FrameWidth

ma:FrameHeight

ma:Locator

owl:sameAs

Róisín Murphy - Pandora

http://www.youtube.com/user/cartoxa

http://www.flickr.com/people/cartoxa/

http://farm3.static.flickr.com/2038/2014900103_d2b12a3c29_m.jpg
How much data is there?

<table>
<thead>
<tr>
<th>Event</th>
<th>Agent</th>
<th>Location</th>
<th>Photos</th>
<th>User</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last.fm</td>
<td>57,258</td>
<td>50,150</td>
<td>16,471</td>
<td>1,425,318</td>
</tr>
<tr>
<td>Upcoming</td>
<td>13,114</td>
<td>0</td>
<td>7,330</td>
<td>347,959</td>
</tr>
<tr>
<td>Eventful</td>
<td>37,647</td>
<td>6,543</td>
<td>14,576</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>108,019</td>
<td>56,693</td>
<td>38,377</td>
<td>1,773,277</td>
</tr>
</tbody>
</table>

1,248,021 geo-tagged photos by propagating information from events!
How fast media are uploaded?
Finding more media that illustrate an event

A. Compute the bounding box area of a venue
B. Retrieve all media geo-tagged in this area
C. Retrieve all media with a similar title
D. Prune the results with visual analysis
E. Extend the result set with all media from the same uploader
A. Bounding box of Nouveau Casino?
B. 74 photos taken in this area this day
C. 85 additional photos with a similar title
D. 6 photos after visual pruning
How is the visual pruning performed?

- Model dataset: photo id + photo geo
- Testing dataset: similar title
- Low-level features used:
  - Color moments, Gabor texture, Edge histogram
- L1 distance on the K-nearest neighbors
- Threshold
  - Min L1 distance between two model image pairs
  - Conservative approach
E. 66 photos after uploader heuristics

cartoixxa

hellerpop

DustGraph / Stefan

13 photos

46 photos
Same process for videos

1 video (id)
3 videos (geo)
26 videos (title)

Visual pruning performed on key frames
Nb positive $> 50\%$
## How illustrated are events?

<table>
<thead>
<tr>
<th></th>
<th>Query By ID</th>
<th>Query By Geo</th>
<th>Query By Title</th>
<th>Visual Pruning</th>
<th>Heuristic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Photos</strong></td>
<td>5</td>
<td>74 (74)</td>
<td>85 (85)</td>
<td>6 (6)</td>
<td>66 (66)</td>
</tr>
<tr>
<td><strong>Videos (title)</strong></td>
<td>1</td>
<td>3 (0)</td>
<td>23 (0)</td>
<td>13 (0)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Videos (title+venue)</strong></td>
<td></td>
<td></td>
<td>10 (10)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 20 events
- Model dataset: 785 photos
- Testing dataset: 1766 photos (1573 positive, 193 negative)
- Results: 439 photos (99% precision, 28% recall)
Generating Visual Summaries
Generating Visual Summaries
Generating Visual Summaries
Generating Visual Summaries
Event Detection

- Detecting events by analyzing user activity on Flickr (uploading pattern)

Accumulated Number of Uploading Photos

Possible Event
Example: the venue Koko

- Ground truth obtained by scraping venue site
  http://scraperwiki.com/profiles/Hou/
Example: the venue Melkweg

- More events detected than event directories listings
- Some events have no illustrative media
Translating the Ontology and the Data
Interfaces

- Perspectives
  - What: Event/Media Centric
  - Who: Social Network Visualization
  - When: Time centric
  - Where: Location Centric
The Back-end

- RDF Repository on a web server with:
  - Sesame2 SPARQL endpoint with a distributed query engine.
  - A RESTful API that provides different methods and JSON representations of resources available in the dataset.
User Interface
User Interface

EventMedia

LIST OF EVENTS

ZZ Top
13 October 2010

High Voltage Festival
24 July 2010 – 25 July 2010

Suikerrock 2010 Kick-off
20 July 2010

Rock in Roma: ZZ Top
12 July 2010

Musilac
July 2010 – 18 July 2010

EventMedia is an Integrative Research Project (IRP) within the PetaMedia Network of Excellence.
Agenda

- A crash course in the world of structured data
  - #microdata, #microformat, #rdfa
  - #rdf, #owl, #skos, #sparql, #linkeddata

- EventMedia (User-centered design approach)
  - LODE: a model for representing events
  - Scraping and interlinking description of events
  - Enriching events with illustrating media
  - Detecting events from social media activities

- Detecting events from human sensing
  - #twitter, #foursquare, #facebook
Citizen Sensors in Action

How Social Media Accelerated Tunisia's Revolution: An Inside View

Social Media Made Tunisian Uprising Possible

http://huff.to/hp0OhA

- Mumbai Terror Attack
- Iran Election 2009
- Haiti Earthquake 2010
- US Healthcare Debate 2009
Citizen Journalism

BBC tells news staff to embrace social media

BBC journalists must keep up with technological change - or leave, the director of BBC Global News Peter Horrocks says

Q&A: Peter Horrocks on the BBC and social media

Images:
http://bit.ly/9GVfPQ,
Business Intelligence

- Trend Spotting, Forecasting, Brand Tracking, Targeted Advertising

- **Sysomos**: Business intelligence by engaging, measuring and understanding activities in Social Media
- **Trendspotting**: Detecting, analyzing and evaluating trends for business.
- **Simplify**: A collaborative platform to monitor, measure and engage customers using Social Media.
- **Shoutlet**: Managing social media marketing communication using a single platform.
- **Reputation.com**: Preserves privacy and defends reputation by protecting attacks on personal information.
What's in a Tweet?

Number of tweets this user has.

Number of favorites this user has.

The timezone and offset (in seconds) for this user.

The user's selected language.

Whether this user is protected or not. If the user is protected, then this tweet is not visible except to "friends".

Number of followers for this user.

DEPRECATED In this context

Whether this user has a verified badge.

DEPRECATED

The place ID

The contributors' (if any) user IDs (http://bit.ly/50npwu).

The URL to fetch a detailed polygon for this place


The type of this place - can be a "neighborhood" or "city"

The country this place is in

The printable names of this place

The contributors enabled

The place


The geo tag on this tweet in GeoJSON (http://bit.ly/6L1Gp).

Whether this user has geo enabled (http://bit.ly/6L1).

Number of users this user is following.

Map of a Twitter Status Object

Raffi Krikorian <raffi@twitter.com>
18 April 2010
Metadata about People

Identification

Network

Activity

Interests
# Metadata about People

<table>
<thead>
<tr>
<th><strong>User Identification Metadata</strong></th>
<th><strong>Interest Metadata</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• User-id</td>
<td>• Author type</td>
</tr>
<tr>
<td>• Screen/Display-name of user</td>
<td>- Trustee/donor, journalist, blogger, scientist etc.</td>
</tr>
<tr>
<td>• Real name of user</td>
<td>• Favorite tweets</td>
</tr>
<tr>
<td>• Location</td>
<td>• Types of lists subscribed</td>
</tr>
<tr>
<td>• Profile Creation Date</td>
<td>• Style of Writing (personality indicator)</td>
</tr>
<tr>
<td>• User description</td>
<td>• No. of Followees</td>
</tr>
<tr>
<td>- Biodata of the user</td>
<td>• Majority of author type of Followees</td>
</tr>
<tr>
<td>- Link to webpage of the user</td>
<td></td>
</tr>
</tbody>
</table>
# Metadata about People

## Activity Metadata

- Age of the profile
- Frequency of posts
- Timestamp of last status
- No. of Posts
- No. of Lists/groups created
- No. of Lists/groups subscribed

## Influence Metadata

(Inferring People Metadata from Network level Information)

- No. of Followers – normal, influential
- No. of Mentions
- No. of Retweets/Forwards
- No. of Replies
- No. of Lists/groups following
- No. of people following back
- Authority & Hub Scores

## Web Presence:

- User affiliations
- KLOUT Score – influence measure  ([http://www.klout.com](http://www.klout.com))
### Metadata about Network

<table>
<thead>
<tr>
<th><strong>Structure Metadata</strong></th>
<th><strong>Relationship Metadata</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Community Size</td>
<td>• Type of Relationship</td>
</tr>
<tr>
<td>• Community growth rate</td>
<td>• Relationship strength</td>
</tr>
<tr>
<td>• Largest Strongly Connected Component size</td>
<td>• User Homophily (based on certain characteristic such as location, interest etc.)</td>
</tr>
<tr>
<td>• Weakly Connected Components &amp; Max(WCC) size</td>
<td>• Reciprocity: mutual relationship</td>
</tr>
<tr>
<td>• Average Degree of Separation</td>
<td>• Active Community/ Ties</td>
</tr>
<tr>
<td>• Clustering Coefficient</td>
<td></td>
</tr>
</tbody>
</table>
## Metadata about Content

<table>
<thead>
<tr>
<th>Explicit Content Metadata</th>
<th>Implicit Content Metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Length, Date/Time, Location of tweet</td>
<td>1. Semantic metadata describing entities (arguably an indirect content-based metadata)</td>
</tr>
<tr>
<td>2. General topic/event</td>
<td>2. Relationships (semantic associations) between entities</td>
</tr>
<tr>
<td>3. Entities mentioned in the tweets {entities includes all real world</td>
<td>3. Sentiments, Perceptions (what people feel about the topic/event),</td>
</tr>
<tr>
<td>entities like person, location, organizations, movies, etc.</td>
<td>4. Intention: seeking information, sharing view-points, intent to buy, spam</td>
</tr>
<tr>
<td>identified using DBPedia, Freebase, GeoNames etc.}</td>
<td></td>
</tr>
<tr>
<td>4. URL mentioned in the tweets</td>
<td></td>
</tr>
<tr>
<td>5. Hashtags used</td>
<td></td>
</tr>
<tr>
<td>6. Sentiment expression</td>
<td></td>
</tr>
</tbody>
</table>

Web-URL mentioned in the tweets (refers to images, videos, blogs, news articles, etc.)
Extracting Entities from Tweets

Name Entity Recognition
Disambiguation

Step 1
- Enter URI

Step 2
- Choose Evaluate Tool

Step 3
- Evaluate Result

AlchemyAPI
Zemanta
DBpedia
Twitris: Semantic Social Web Mash-up

Facilitates understanding of multi-dimensional social perceptions over SMS, Tweets, multimedia Web content, electronic news media

Searching on Twitter

See what's happening — right now.

eindhoven entertainment event

Trending topics: #itmakesmesmilewhen, #scholes, #singleladies, LULUS RAME-RAME, NAIK KELAS RAME-RAME, Serious Question, Daesung, Jeremy Kyle, Sean Kingston, IKEA

Twitter Home · About Twitter Search · API · Jobs · Install Search Plugin

© 2011 Twitter, Inc.
Issues with Multiple Keywords Search

Results for *eindhoven entertainment event*

No results for *eindhoven entertainment event*
Let’s try to search with One Keyword

See what's happening — right now.

eindhoven

Trending topics: #itmakesmesmilewhen, #scholes, #singleladies, Serious Question,
NAIK KELAS RAME-RAME, LULUS RAME-RAME, Daesung, Sean Kingston, Jeremy Kyle, IKEA

Twitter Home · About Twitter Search · API · Jobs · Install Search Plugin

© 2011 Twitter, Inc.
8 minutes ago via HANDELSPIRJZEN.nl - Reply · View Tweet

Jikke073: @MrrMayk oo viel da zo erg tege dan ?? k ga daar miss ook heen na die school anders eindhoven :P
8 minutes ago via Twitter · Reply · View Tweet · Show Conversation

valkeindhoven: @Robert043 Dat klopt! Bijna alle hotels in Eindhoven zitten vol, i.v.m. concert Guus Meeuwis.
8 minutes ago via web · Reply · View Tweet · Show Conversation

poeetweet: RT @Kailimero: Geen bom gevonden in IKEA Eindhoven. Wel 364395 kwijtgeraakte schroeven, 6 vermiste kinderen en een kabouter-lijkje in verregaande staat.
8 minutes ago via web · Reply · View Tweet

josmeesse: Luister volledig af naar het woord van de bevelhebber! Mee en vanaf vandaag is er spraak van strijdDeclarencie! https://t.co/0w7G1B
8 minutes ago via Twitter · Reply · View Tweet

izzingsa: Luister volledig af naar het woord van de bevelhebber! Mee en vanaf vandaag is er spraak van strijdDeclarencie! https://t.co/0w7G1B
9 minutes ago via Twitter · Reply · View Tweet

StanLenssen: I'm at Vrije Werkplaats Stan Lenssen (Toorenlaan 26-36, Beukenlaan, Eindhoven) http://4sq.com/jSYVDD
9 minutes ago via foursquare from Eindhoven, Brabant · Reply · View Tweet
Prinsesabrina: Heb je vakantie, moet je alsnog vroeg op omdat je neef wil afspreken helemaal in Eindhoven O_o!
14 minutes ago via ÜberSocial • Reply • View Tweet

RobPapen: This evening in Eindhoven. 20.30hr Demo of Punch. http://blog.i4muzique.nl/2011/05/rob-papen-komt-met-punch-naar-i4-muzique/
15 minutes ago via web • Reply • View Tweet

16 minutes ago via web • Reply • View Tweet • Show Conversation

Fixview: Quality Manager in de regio Eindhoven: http://lnkd.in/WCvCRn
16 minutes ago via LinkedIn • Reply • View Tweet

Vooorscht: School op het oude philips terrein naar een andere locatie. Daarna door naar Waddinxveen aan het buitengebouw.
16 minutes ago via LinkedIn • Reply • Show Conversation

StephenvdB: #Eindhoven...Morgen naar Utecht, kunnen we het goed doen...
17 minutes ago via web • Reply • View Tweet • Show Conversation

Esther kersten: @ Maartje Kersten Tja, Utrecht verklekt natuurlijk naast wereld(licht)stad #Eindhoven, dus maak ik er het beste van. Maar regenen kunnen we wel!
17 minutes ago via web • Reply • View Tweet • Show Conversation
Next Saturday @thatsimpsonsonguy aka Guilty Simpson will be performing at Area51 in my hometown Eindhoven. #realliveshit #iwillspinrecords
Relation Discovery Framework

Entity extraction & semantic enrichment

Relation discovery

Relation type

temporal constraints

weighting scheme

source selection

Applications
- Browsing support
- Query suggestions
- Schema enrichment

Ilknur Celik, Fabian Abel, Geert-Jan Houben

Web Information Systems, TU Delft
Julian Assange, the founder of WikiLeaks, is under arrest in London...
Relation Learning Strategies

- **Relation:**
  
  \[ \text{relation}(e_1, e_2, \text{type}, t_{\text{start}}, t_{\text{end}}, \text{weight}) \]

- **Relation Learning strategy:**
  - **Input:** entity \( e_1 \) and \( e_2 \), time period \((t_{\text{start}}, t_{\text{end}})\)
  - **Challenge:** infer weight and type of the relation for the given

- **Weighting according to co-occurrence frequency:**
  - **Tweet-based:** count co-occurrence in tweets
  - **News-based:** count co-occurrence in news
  - **Tweet-News-based:** count co-occurrence in both tweets and news
Where do relationships emerge faster?

![Graph showing time difference (in days) of first occurrence of relationship]

- Person-Movie
- Person-MusicAlbum
- Product-OperatingSystem
- Person-PoliticalEvent
- Company-Product
- Person-City
- Company-City

- News is faster
- Twitter is faster

Speed of strategies is domain-dependent

http://wis.ewi.tudelft.nl/icwe2011/relation-learning/
On Conferences … we Tweet
Rich Activity Twitter Event Data

- Take Twitter archives from TwapperKeeper
- Enrich Tweets with relevant DBPedia concepts using Zemanta
- Rely on existing Linked Data about talks to perform the mappings.

Milan Stankovic & Matthew Rowe: *Mapping Tweets to Conference Talks*, SDOW 2010
Find the correspondence

baojie Jie Bao
Do you know each page on #iswc2010 annotated? e.g., the program http://bit.ly own app! #metadata
from MIT, Cambridge
3 hours ago

FlyingSourcr Flying Sourcer
3 hours ago

baojie Jie Bao
ISWC Mobile: report bug to @alvaro@grao #metadta http://goo.gl/ByJy
from MIT, Cambridge
4 hours ago

baojie Jie Bao
ISWC Mobile update. Photo & Contact c and time of each talk; Better Browsing. #

Research Track

- Using Reformulation Trees to Optimize Query Answering
  Authors: Yingjie Li, Jeff Hefflin

- SPARQL Query Optimization on Top of DHF
  Authors: Zoi Kaoudi, Kostis Kyzirakis, Mancini

- Summary Models for Routing Keywords to Data
  Authors: Thanh Tran, Lei Zhang, Rudi Studer

- Linked Data Query Processing Strategies
  Authors: Gunter Ladwig, Duc Thanh Tran

- When owl:sameAs isn't the Same: An Analysis
  Authors: Harry Halpin, Patrick J. Hayes, James P. McColl Thompson

- SameAs Networks and Beyond: Analyzing owl:sameAs in Linked Data
  Authors: Li Ding, Joshua Shinavier, Zhenning Shi

- SAOR: Template Rule Optimisations for D1 Triples
  Authors: Aidan Hogan, Jeff Z. Pan, Axel Polleres

- Signal/Collect: Graph Algorithms for the (S3) Web
  Authors: Philip Stutz, Abraham Bernstein, Wilfred Schulte

- Finding the Achilles Heel of the Web of Data
meets

Search | Photos | Groups | People

Everyone's Uploads | foursquare venue=

Sort: Relevant | Recent | Interesting

View: Small | Medium | Detail | Slideshow

29/06/2011 - Lecture at the 2nd Summer School on Social Media Retrieval (S3MR) - Antalya, Turkey
Gundotra: “We believe online sharing is broken. And even awkward. We think connecting with other people is a basic human need. We do it all the time in real life, but our online tools are rigid. They force us into buckets — or into being completely public”
Conclusions

- The importance of structured data

- EventMedia
  - Dataset part of the Semantic Web
  - LODE used by the UK Archives Hub
  - Method for finding media illustrating *scheduled* events
  - Method for detecting events from social media
  - Social Event Detection Task

- Event-based approach for users to explore, annotate and share media
  - UX can help semantics, semantics can help UX
  - Outstanding challenges in interlinking and curating the data
Credits

- **EURECOM**: Houda Khrouf, Giuseppe Rizzo
- **CWI**: Ryan Shaw, André Fialho, Lynda Hardman
- **Google/Yahoo!**: Thomas Steiner, Peter Mika
- **Colleagues**: Fabien Gandon, Alexandre Passant, Amit Sheth, Fabian Abel, Milan Stankovic, Matthew Rowe
- … and the “media sharers”