Data Enrichment Service (DES)

*Linking data from unstructured and semi-structured sources*
About TSO

- Long established profitable UK firm
- 365 employees
- Global reach
- Semantic solutions team
- OpenUp® Platform used to harvest, enrich, store and publish data
OpenUp® TSO’s semantic platform

- **Harvest**: Aggregation of data from web, APIs, databases and files
- **Enrich**: Extracting useful data and converting to re-usable formats
- **Store**: Highly scalable database storage and query engine
- **Publish**: Websites and APIs to reach data users

Automated processes that deliver reliable data
Data Enrichment Service (DES) and semantic enrichment

The automatic extraction of information from large amounts of unstructured content

What is it for?

Automated tagging of documents
Creation of new information outputs
Improve searching and relevance of online products
Linking related information
Automated classification

Bootstrap Linked Data by automatically generating lots of it
Welcome to the OpenUp Data Enrichment Service

The DES is infrastructure for hosting information extraction processes. A core extraction pipeline is available that provides basic annotations that many processes will need. This can be used to bootstrap a more domain-specific extraction process. DES uses GATE underneath and any GATE application can potentially be run on DES.

This site allows you to try out the core DES extraction process. Feel free to have a play ...

For developers more information be found in the Developer Zone.

Try Some Text ...

William Hague launches full 'audit' of EU law and the UK
Comments (276)
William Hague at the 2011 Conservative party conference Mr Hague is under pressure from his party to repatriate powers from the EU
Continue reading the main story
Related Stories

Case for EU vote powerful - Hague
PM 'prepared to consider EU vote'
EU vote backed by 100 Tony MPs

The government is launching a "comprehensive audit" of European Union powers and their impact on the UK.
Results of Processing

Below are the results of processing your text, along with a key indicating how to interpret the data and a full list of the extracted information.

<table>
<thead>
<tr>
<th>Output For Your Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>William Hague launches full ‘audit’ of EU law and the UK</td>
</tr>
<tr>
<td>Comments (276)</td>
</tr>
<tr>
<td>William Hague at the 2011 Conservative party conference Mr Hague is under pressure from his party to repatriate powers from the EU</td>
</tr>
<tr>
<td>Continue reading the main story</td>
</tr>
<tr>
<td>Related Stones</td>
</tr>
<tr>
<td>Case for EU vote powerful - Hague</td>
</tr>
<tr>
<td>PM 'prepared to consider EU vote'</td>
</tr>
<tr>
<td>EU vote backed by 100 Tory MPs</td>
</tr>
<tr>
<td>The government is launching a &quot;comprehensive audit&quot; of European Union powers and their impact on the UK</td>
</tr>
<tr>
<td>In a Commons statement, Foreign Secretary William Hague outlined plans to scrutinise every aspect of EU law.</td>
</tr>
<tr>
<td>The audits is a &quot;necessary and positive part of reforming Europe&quot;, he said, and is due to conclude in 2014</td>
</tr>
<tr>
<td>Hague said that an urgent strategy was needed sooner than 2014 to inform the EU’s response to greater political integration in the EU</td>
</tr>
<tr>
<td>Shadow foreign secretary Douglas Alexander also sought to highlight divisions in the coalition over the subject, pointing out that former Lib Dem leader Charles Kennedy was the only backbencher from his party to have listened to the statement in person.</td>
</tr>
<tr>
<td>Continue reading the main story</td>
</tr>
</tbody>
</table>

*Start Quote*

<table>
<thead>
<tr>
<th>Key</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continent</td>
</tr>
<tr>
<td>Position</td>
</tr>
</tbody>
</table>

Note that not all the above annotations may be listed below for conciseness (marked *), however they will still display in the left pane.

<table>
<thead>
<tr>
<th>Classification (Alpha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category: <a href="http://des.tso.co.uk/core/2012">http://des.tso.co.uk/core/2012</a></td>
</tr>
<tr>
<td>Probability: 0.3600491</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Extracted Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continent</td>
</tr>
<tr>
<td>Europe</td>
</tr>
<tr>
<td>Country</td>
</tr>
<tr>
<td>UK</td>
</tr>
<tr>
<td>Date</td>
</tr>
<tr>
<td>2011</td>
</tr>
<tr>
<td>2014</td>
</tr>
<tr>
<td>Organisation</td>
</tr>
</tbody>
</table>
London 2012: Olympians and Paralympians cheered by crowds

Chris Hoy

Details
Date of Birth: 1976-03-23

Abstract
Sir Christopher Andrew "Chris" Hoy, MBE is a British and Scottish track cyclist representing Great Britain and Scotland. He is a multiple world champion and Olympic Games gold medal winner. With his three gold medals in Beijing 2008, Hoy became Scotland's most successful Olympian, the first Briton to win three gold medals in a single Olympic games since Henry Taylor, in 1908, and the most successful Olympic male cyclist of all time.

Wikipedia Page
http://en.wikipedia.org/wiki/Chris_Hoy
The chart above shows the number of mentions each week for the past six weeks for this person. Click a bar to view that week.

Total mentions: 25, Overall rank: 9

Documents for the Week Starting 25-09-2012
- Terror suspect fights extradition 01-10-2012
- Archbishop move ‘may take month’ 20-09-2012
- New portrait of Queen is unveiled 20-09-2012
- Ministers take steps to strengthen the Commonwealth 20-09-2012
- Military awards honour battled bravery 28-09-2012
Infrastructure for extraction processes

Data sources:
- Wikipedia
- data.gov.uk
- Ordnance Survey
- GeoNames
- London Gazette and others...

Entities extracted:
- Company names
- Organisation names
- Towns
- Countries
- Dates
- Facilities
- Events and others...

DES

Starter DES (generic) annotator
More than a ‘string matching’ engine

DES needs to disambiguate what it finds:

Paris Hilton: the hotel in Paris or the heiress?

Orange: the colour, the fruit, or the phone company?

Boston: UK or Massachusetts?
Extendable to other domains

For domain specific annotations DES can use:

- Training
- Semi-controlled language
- Thesaurus
- Controlled vocabulary

Legislation is semi-controlled
Citation, extent and commencement

Cited as: http://www.legislation.gov.uk/idl/kelt/2011/948

1.—(1) This Order may be cited as the Council Tax (Discount Disregards) (Amendment) Order 2011.

Come into force
This Order
Date: 13th May 2011

Extent
This Order
For the countries
England and Wales

(2) This Order extends to England and Wales and comes into force on 13th May 2011.

Amendments to the Council Tax (Discount Disregards) Order 1992

2. The Council Tax (Discount Disregards) Order 1992(2) is amended as follows.

Amendments to Article 5 relating to prescribed educational establishments

Reference: s.5, (Text is 'article 5')
For
Reference: paragraph 2,(Text is 'paragraph (2)')
Substitute
the following paragraph
The amendment

2 For the purposes of paragraph 5(4) of Schedule 1 to the Act, the bodies prescribed for the definition of "institution" are those institutions situated in England and Wales which—a in relation to students, are defined as prescribed educational establishments in Part I of Schedule 2 to this Order; and b in relation to student nurses, are defined as prescribed educational establishments in Part II of that Schedule.

3. In article 5 (prescribed educational establishments), for paragraph (2) substitute the following paragraph—
Semantic enrichment solutions

- Built on GATE Open Source technology (developed by Sheffield University)
- GATE is used by major organisations:
  - Platform as a Service (PaaS)
  - Extracts information from variety of forms: Word, XML, RDF, TXT, PDF
TSO’s SPARQL editor Flint is open source
‘Opened up’ bibliographic data

- Datasets available for public access as ‘Linked Data’
- Datasets include bibliographic data about articles published 1869 to the present day
- Available under a Creative Commons Zero dedication (you can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission)
- Access to datasets published as Linked Data and made available through SPARQL services
- Queryable through a service endpoint
- TSO hosts the RDF triples in our triplestore
- Just released another 250m RDF triples

‘… Linked data is an important step in the evolution of scientific publishing and the launch of the platform has been met with a positive response from the scientific community. We are looking forward to continuing to work with TSO as we further enhance the Linked Data Platform.’ Head of Digital First Amanda Ward
Contacts

http://openup.tso.co.uk/des
http://openup.tso.co.uk/des/enriched
http://openup.tso.co.uk/developer/des/api
Confidentiality statement

The contents of this document together with all other information, data, materials, specifications or other related documents provided by Williams Lea ("WL") (together “materials”) shall be treated at all times by the recipient as the confidential and proprietary information of WL. The recipient shall not disclose any such materials to any third parties without the express, prior written approval of WL. Where such express approval is granted by WL, the recipient shall ensure that all third parties to whom disclosure is made shall keep any such materials confidential and shall not disclose them or any part of them to any other person. All intellectual property rights in the materials shall remain the property of WL, or its third party licensors, and are protected by copyright.

© 2012 Williams Lea Group

Disclaimer

This document may be incomplete without reference to any oral briefing provided by WL, reflects current conditions and WL’s views as of this date and is subject to correction or change at any time. Although the information contained in this document is believed to be accurate in all material respects, neither WL nor any of WL's advisers, agents, officers or employees accepts responsibility or liability for or makes any promise, representation, statement or expression of opinion or warranty, express or implied, with respect to the accuracy or completeness of the content of this document (to the extent permissible by law) unless and save to the extent that such promise, representation, statement or expression of opinion or warranty is later expressly incorporated into a legally binding contract.