UNLEASH THE TRIPLE:

LEVERAGING A CORPORATE DISCOVERY INTERFACE

THE OECD CASE

Mary-Ann Grosset, Digital Practice Team Manager
Thierry Vebr, Coordinator Information Systems
Jan-Anno Schuur, Coordinator Information Specialist

12 September 2017

OECD
BETTER POLICIES FOR BETTER LIVES

OECD.org | Data | Publications | More sites | News | Job vacancies

OECD
BETTER POLICIES FOR BETTER LIVES

www.oecd.org

> A to Z
Search oecd.org

OECD Home | About | Countries | Topics | > Français

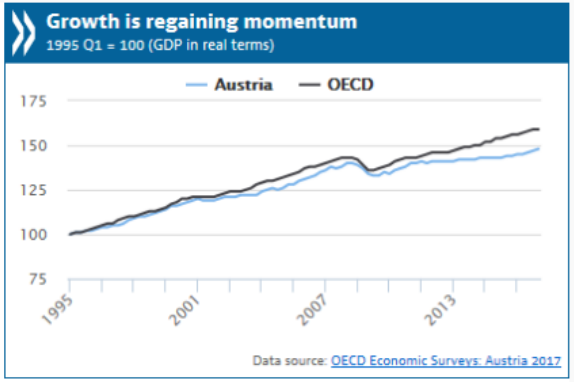
What's new

» Economy: Digitalisation can help Austria boost growth and inclusiveness

» Government: Decline in public investment should be addressed as costs stabilise

» G20: Secretary-General Gurría reaffirms need for global co-operation amid progress at summit

» Agriculture: Slow growth in demand expected to keep world food prices low



17 July 2017

The Austrian economy is strengthening, supported by recent tax reforms and a pick-up in international trade. With confidence rising and the

Focus

Stay informed with myOECD

myOECD

Want the latest OECD news and statistics delivered straight to your inbox? Click to connect!

» Read more

35 member countries

Mission: promote policies that will improve the economic and social well-being of people around the world

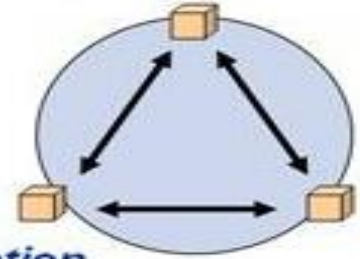
OECD
Organisation for Economic Co-operation and Development

Who drives the OECD's work?

Council

Oversight and strategic direction

Representatives of member countries and of the EuropeanCommission; chaired by the Secretary-General; decisions taken by consensus



Committees

Discussion and implementation

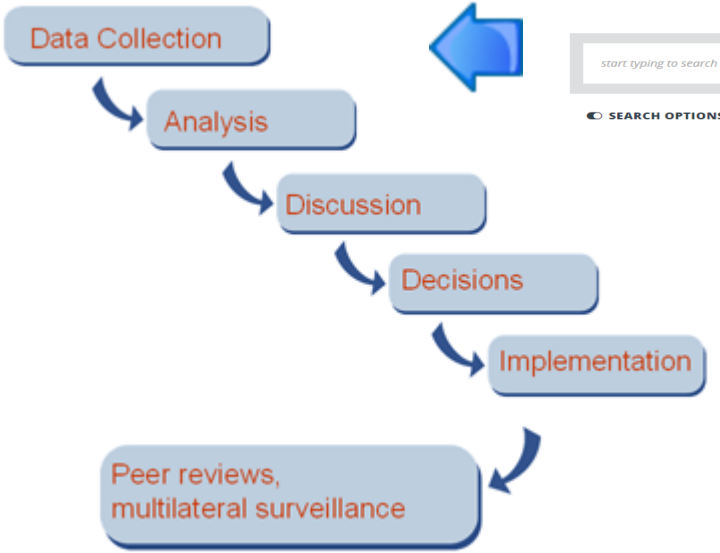
Representatives of member countries and of countries with Observer status work with the OECD Secretariat on specific issues

Secretariat

Analysis and proposals

Secretary-General
Deputy Secretaries-General
Directorates

OECD's way of working



O.N.E Sight

Reading Assistant
Less time to identify and collect information
More time available to analyse

Internal and external sources
No language or word constraints

Makes suggestions to:
- Narrow down search
- Discover new areas

Fully semantic environment
Content and queries tagged using semantic robots using taxonomies/ontologies and linguistic patterns
Tags stored in RDF to link content

O.N.E Sight

Resources	11-Jul-17	
	Number of resources	Semantic annotations
Central taxonomies	10,234	n/a
Domain-specific taxonomies	29,081	n/a
Indicators	8,245	944,279
Models and ontologies	1,463	n/a
Internal resources	295,637	4,614,464
Datasets	1,281	108,930
External resources	4,703,396	40,398,473
Surveys	44,405	87,140
Total	**5,093,742**	**46,153,286**

How did we get to where we are

- What:
  - Put structure into unstructured information
  - Extract knowledge from content
  - Enable repurposing of knowledge and content nuggets
- How:
  - Fragmenting content (information)
  - Storing content as structured XML
  - using taxonomies, semantic tagging
  - Storing taxonomies, tags, links and triples
  - Using linked data capabilities
- Deliverables:
  - PoC to search and discover content (October 2015)
  - Complex semantic analysis tools identifying indicators, trends,…. (from April 2015 to current)
  - Project specific reading and tagging assistants (from May 2016 to current)
  - O.N.E Sight Reading assistant (April 2017)

Phase 1 : Capacity building
September 2014 to October 2015

- Move from Librarians/Records Managers to Taxonomy , semantic and linked data experts
- Training/workshops in using applications and creating services
- Practical hands-on experience to create and maintain
- Stakeholder complex semantic use cases
- Agile methodology

And a lot of

Creativity
Innovation
Curiosity
Motivation

Dedication
Team work
Collaboration
Enthusiasm

Phase 2: Building the foundations
January to October 2015

- Corporate taxonomies
  - Topics
  - Geographical areas
- Semantic robots to
  - Identify candidate terms to update the corporate taxonomies
  - Identify topics/geographical areas used in information at document level
  - Identify topics/geographical areas used at fragment level
- Modelise content
  - Flexible , adaptable
  - Easy to expand
  - Data source Independent
- Store all content as RDF
- Develop PoC

Phase 3: Stakeholder semantic applications November 2015 to today

- Analyse sections of reports using a domain specific ontology and develop a service to allow manual tagging of sections, Display results in context
- Analyse responses to questionnaires at fragment level using central topics and geographical area ontologies; identify trends and present results in context
- Analyse legislation from 4 different countries (step 1) in French, English and Spanish to identify best topics and/or all topics, depending on the desired view
- Analyse press releases from Reuters and AFP to identify acts of civil unrest in African countries; develop a service to allow manual validation of indicators following their identification through complex semantic tagging
- Enrich articles from Elsevier, and RSS feeds to present targeted alerts to stakeholders
- Develop a products taxonomy to semantically tag product Recalls

Phase 4: O.N.E Sight Digital strategy project November 2016 to today (April 2017 for V1 release)

- Build on experience:
  - Creating taxonomies/ontologies
  - Creating diverse and complex semantic robots and web-services
  - Creating complex SPARQL queries
  - Developing easy to use solutions to meet complex stakeholder semantic analysis needs
- Develop a fully semantic and linked data environment
- One-stop shop for policy analysts to conduct research
- Semantically enrich content from internal and external sources

The O.N.E Sight Team

Knowledge Management Unit

Development and maintenance
Model
SPARQL queries
Knowledge contextualisation
Maintenance of environments
(taxonomy, semantic),
triple and XML stores

Taxonomy
Semantic robots
Corpora
Validation
Disambiguation

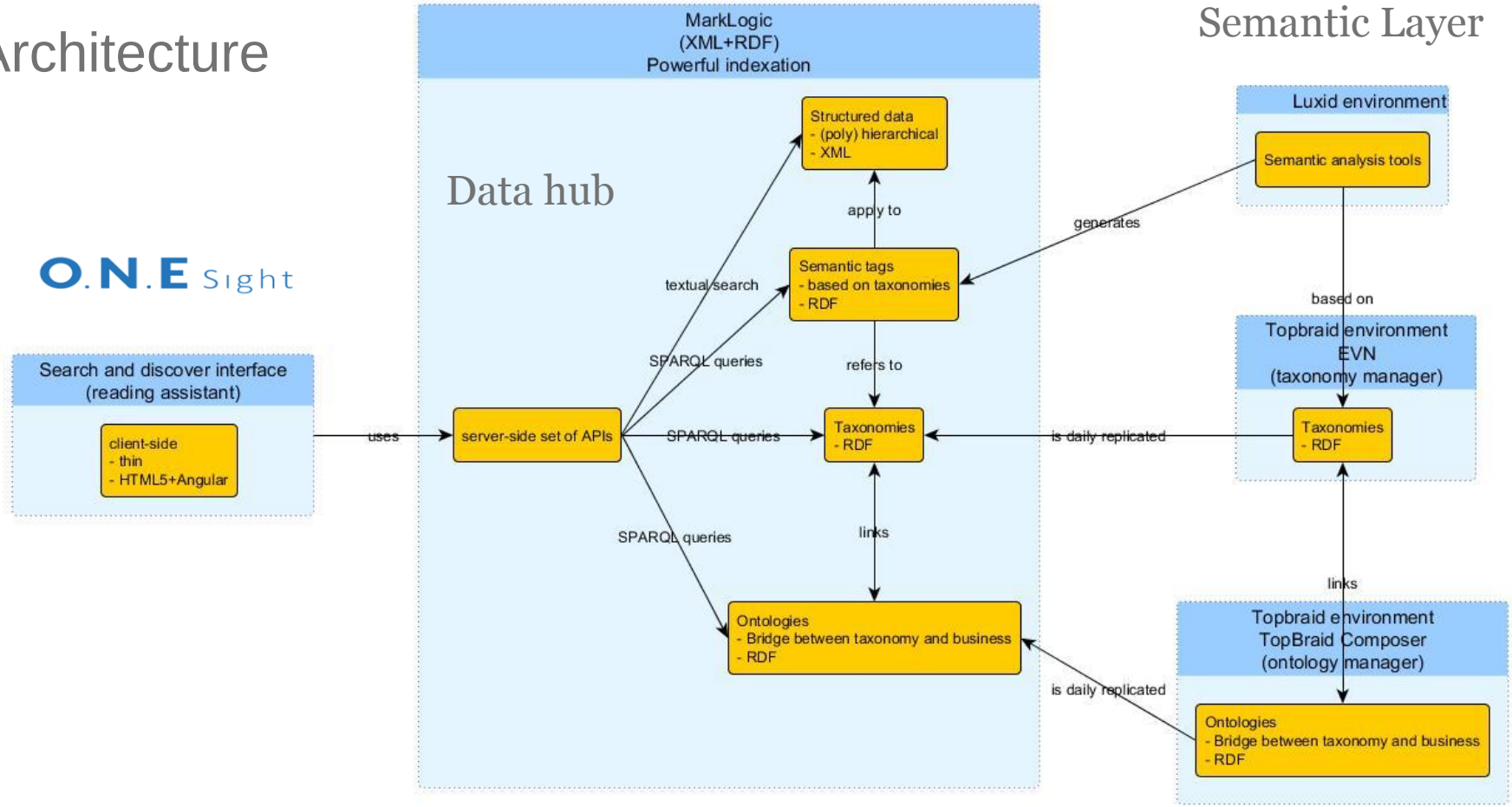
Key stakeholders

Functionality
Domain specific taxonomies
Semantic requirements

External developers

UX design
Interface development

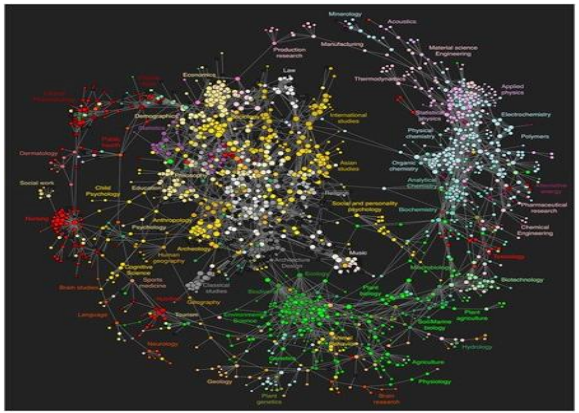
Architecture



Breaking the barriers

Taxonomies/Ontologies

Of LANGUAGE and WORDS



Create concepts and link them so that machines can use them as a cloud of knowledge (Semantic layer)

Capture and formalise Subject-Matter-Experts' knowledge

Identify synonyms, related terms in official languages

Use URIs not labels

Use in:

- Semantic analysis tools
- Search and discover tools

Breaking the barriers

cows

vaches

http://kim.oecd.org/Topics#T4658

Cheese fondue

Fondue au fromage

http://kim.oecd.org/Topics#T127

produced by

milk

lait

http://kim.oecd.org/Topics#T665

ingredient of

mice

souris

http://kim.oecd.org/Topics#T7520

derived from

likes

hates

subtype of

gruyère

gruyère

http://kim.oecd.org/Topics#T909

cheese

fromage

http://kim.oecd.org/Topics#T222

cats

chats

matous

http://kim.oecd.org/Topics#T7520

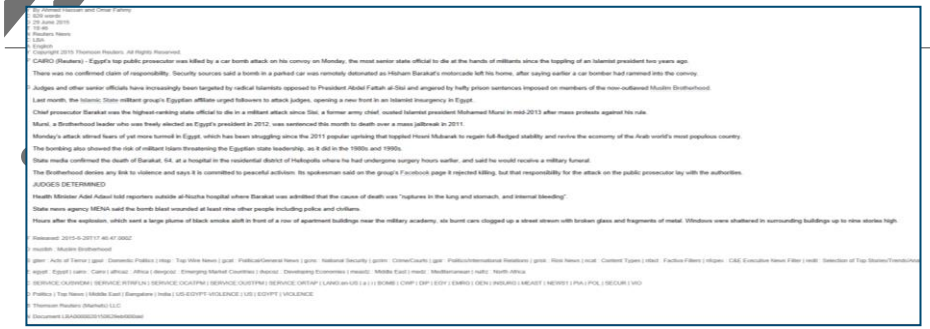
SEMANTICS
FUN AND GAMES

How do we create the semantic robots?

- Use lexicon (taxonomies)
- Use text patterns (part of speech)
- Identify relationships (ontologies)
- Test the results on a set of documents
- Debug - Disambiguate
- Test on the complete corpus
- Put in production using Web Services

36 different robots in production

Semantic enrichment



Governance Indicators – Jet Overview Engine annotations as from 2016-02-03

Selected Indicators | Actors | Events | Objects | Terrorist Group | Sectors

2015 | Egypt [1940] | Indicator-07 - Violence resulting in death [816]

Car bomb attack kills Egypt's top public prosecutor close

CAIRO (Reuters) - Egypt's top public prosecutor was killed by a car bomb attack on his convoy on Monday, the most senior state official to die at the hands of militants since the toppling of an Islamist president two years ago.

There was no confirmed claim of responsibility.

Security sources said a bomb in a parked car was remotely detonated as Hisham Barakat's motorcade left his home, after saying earlier a car bomber had rammed into the convoy.

Judges and other senior officials have increasingly been targeted by radical Islamists opposed to President Abdel Fattah al-Sisi and angered by hefty prison sentences imposed on members of the now-outlawed Muslim Brotherhood.

Last month, the Islamic State militant group's Egyptian affiliate urged followers to attack judges, opening a new front in an Islamist insurgency in Egypt..Chief prosecutor Barakat was the highest-ranking state official to die in a militant attack since Sisi, a former army chief, ousted Islamist president Mohamed Mursi in mid-2013 after mass protests against his rule.

Mursi, a Brotherhood leader who was freely elected as Egypt's president in 2012, was sentenced this month to death over a mass jailbreak in 2011.

Monday's attack stirred fears of yet more turmoil in Egypt, which has been struggling since the 2011 popular uprising that toppled Hosni Mubarak to regain full-fledged stability and revive the economy of the Arab world's most populous country.

The bombing also showed the risk of militant Islam threatening the Egyptian state leadership, as it did in the 1980s and 1990s.

State media confirmed the death of Barakat, 64, at a hospital in the residential district of Heliopolis where he had undergone surgery hours earlier, and said he would receive a military funeral.

The Brotherhood denies any link to violence and says it is committed to peaceful activism.

Its spokesman said on the group's Facebook page it rejected killing, but that responsibility for the attack on the public prosecutor lay with the authorities.

JUDGES DETERMINEDHealth Minister Adel Adawi told reporters outside al-Nozha hospital where Barakat was admitted that the cause of death was "ruptures in the lung and stomach, and internal bleeding".

State news agency MENA said the bomb blast wounded at least nine other people including police and civilians.

Hours after the explosion, which sent a large plume of black smoke aloft in front of a row of apartment buildings near the military academy, six burnt cars clogged up a street strewn with broken glass and fragments of metal.

Windows were shattered in surrounding buildings up to nine stories high."It felt like the apartment was lifted up and dropped right back on the ground.

When I felt it I said to my family, this is it, death has come for us," said 17-year-old Khaled Youssef.

Sisi's office mourned Barakat's death and canceled celebrations prepared for Tuesday to mark the anniversary of mass unrest that led to Mursi's fall, and said the perpetrators would draw "the most severe punishment".

The government, however, declared June 30 a national holiday in commemoration of the 2013 protests, state media said. "These kinds of vicious attacks will not deter the state from continuing its path of development, the adoption of rights, and realizing the hopes and aspirations of the Egyptian people," the presidency's office said in a statement according to MENA. The Egyptian judiciary says it is independent of the government and military. But some of Egypt's judges have drawn accusations of blatant bias by handing down lengthy jail terms and mass death sentences against Islamists.

"Terrorism killed the top man of our prosecution but despite this we will not be scared and we will continue our work," said Judge Ashraf Abdelhady as he left the hospital. A presidential spokesman told

```
<macro name="in36mc3">
  <!-- Government Actors -->
  <e>^@Entity^@Actor^@AssembléeNationale</e>
  <e>^@Entity^@Actor^@ChefDeLArmee</e>
  <e>^@Entity^@Actor^@ChefDuGouvernement</e>
  <e>^@Entity^@Actor^@Consulat</e>
  <e>^@Entity^@Actor^@ForcesSécurité</e>
  <e>^@Entity^@Actor^@Gouvernement</e>
  <e>^@Entity^@Actor^@GroupesArmés</e>
  <e>^@Entity^@Object^@Législation</e>
  <e>^@Entity^@Actor^@Magistrat</e>
  <e>^@Entity^@Actor^@Maire</e>
  <e>^@Entity^@Actor^@Mairie</e>
  <e>^@Entity^@Actor^@Municipal</e>
  <e>^@Entity^@Sector^@Nationale</e>
  <e>^@Entity^@Actor^@NouveauChefDuGouvernement</e>
  <e>^@Entity^@Actor^@NouveauGouvernement</e>
  <e>^@Entity^@Actor^@Officier</e>
```

Search for protests in Africa

Returns articles covering protests, demonstrations, manifestations, etc...

Identifying knowledge nuggets

Indicator 1 – Strike in the Public sector

Must contain

Sector: Public

Event: Strike

Sector: Public

Event: Strike

Indicator 1 – Strike in the Public sector

Recall, Precision and F-Measure

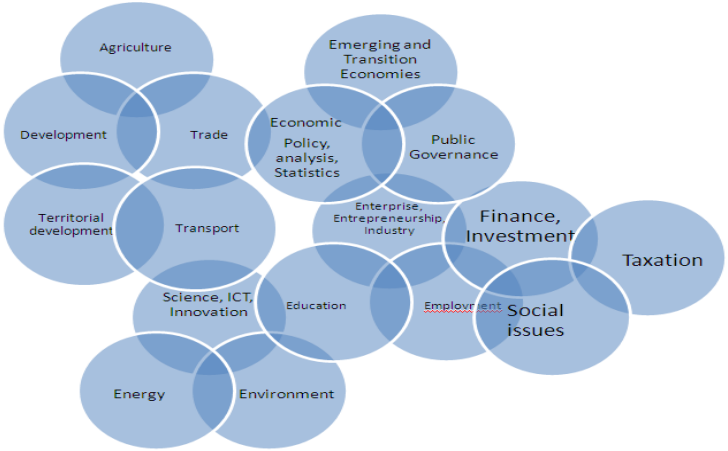
Average rate 95% - Less not acceptable

How do we achieve this:

- Extensive work on taxonomies
- Selection of Corpora
- Validation
- Disambiguation (limit fuzziness, provide context, specify acronyms…)

Golden Corpora and Validation

Corpus (group) of documents used to validate and check the quality of the skill cartridges used to semantically enrich information



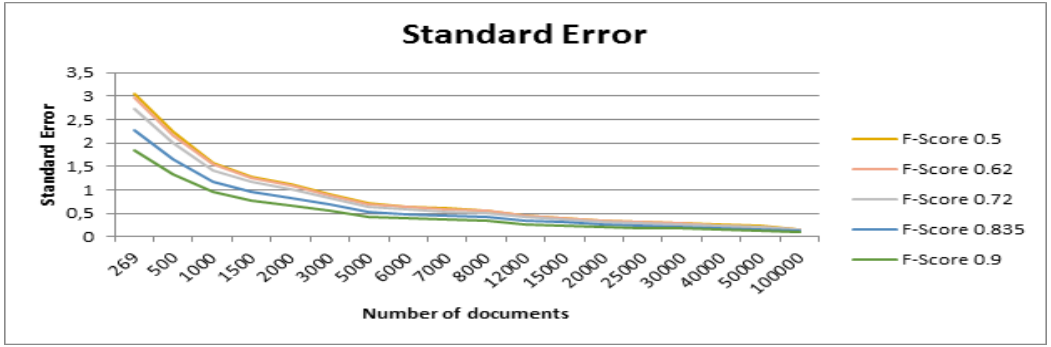
The OECD Golden Corpora contain

- OECD Publications
- Official Documents
- 17 themes
- both official languages
- items manually tagged against the central taxonomy

Golden Corpora :Size and Quality

- 1.223 publications
- Error margin of 0,9.

Provide a realistic overview of the content and volume covered in the different domains.



Theme	Total	XML	En	Fr	% XML
Agriculture	45	18	9	9	40
Development Cooperation	89	36	18	18	40
Economic Policy	176	163	84	79	93
Education	68	22	16	6	32
Employment	42	22	11	11	52
Energy	76	4	2	2	5
Enterprise Entrepreneurship	76	44	30	14	58
Environment	88	56	30	26	64
Finance Investment	167	111	58	53	66
Public Governance	94	66	38	28	70
Science	56	32	16	16	57
Social issues	70	52	27	25	74
Taxation	37	23	15	8	62
Territorial Development	37	28	15	13	76
Trade	61	19	12	7	31
Transport	41	8	3	5	20
Total	1223	704	384	320	58

OECD index for cartridge reference created with the XML publications.

The index provides the basis of OECD language that is used by the cartridge to rank extraction results.

Disambiguation

Attack owl returns in Oregon, targeting government workers

PORTLAND, ORE. | BY COURTNEY SHERWOOD

An owl that achieved notoriety last winter for attacking joggers in an Oregon park has returned and is now turning its talons on government workers, state officials said on Wednesday.

The barred owl has clawed at least three people outside the state Capitol in Salem in a series of attacks since late November, city parks department spokeswoman Tibby Larson said.

Scientists investigate suspected meteorite death in Tamil Nadu

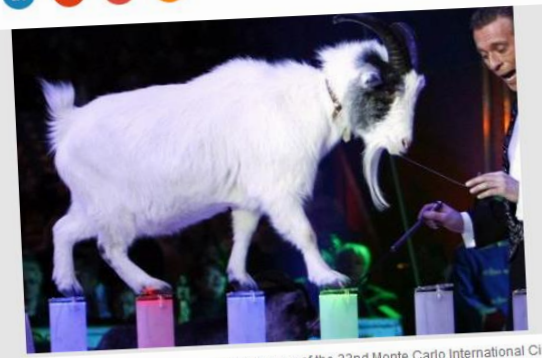
CHENNAI | BY SANDHYA RAVISHANKAR

Indian scientists are investigating whether a man was killed by a meteorite, which if confirmed would be the first recorded death from falling fragments of space rock in almost 200 years.

Jayalalithaa Jayaram, the chief minister of Tamil Nadu, has said a bus driver at a college in her state was killed by the meteorite and awarded 100,000 rupees ($1,470) in compensation to his family.

Goat detained over armed robbery

LAGOS



A magician performs with a goat during the opening ceremony of the 33nd Monte Carlo International Circus Festival in Monaco January 15, 2009.
REUTERS/ERIC GAILLARD

Police in Nigeria are holding a goat on suspicion of attempted armed robbery.

Vigilantes took the black and white beast to the police saying it was an armed robber who had used black magic to transform himself into a goat to escape arrest after trying to steal

Disambiguation is key & takes a lot of time

It’s raining cats and dogs

Queen Elisabeth

Multi-view annotation graphs

We use several semantic robots,

based on several different taxonomies

(generic, innovation-oriented, etc…)

We tag a same resource in different ways

We can see a same resource with different « semantic » angles of view

Demonstration

Outreach

- 5-year programme
- Looking for candidates from OECD member countries for 6 month internships, or short term assignments
- To work with us on specific semantic projects
- Can be associated to a PhD thesis
- 6 to 18K€ funding envelope per project

Contact

For additional information, internship possibilities to work on semantic projects :

mary-ann.grosset@oecd.org
Knowledge Management Unit
Digital Knowledge and Information Services
2 rue André Pascal
75775 Paris cédex 16
France
Tel: +331 45 24 81 53