



# OpenCalais

Issa Mahasneh  
servit Open Source Solutions

# Calais: What is it?

- Have you ever heard of “Semantic Web”?
- Why not applying natural language processing to “content tags”?
- Is there a way to identify important metadata automatically in some text?

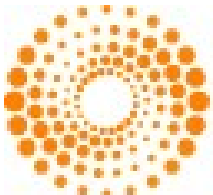
# Calais Will Help You!

- An initiative to encourage the wide deployment of **semantic technologies** in the information and content marketplaces.
- Is a set of services that can identify and extract “entities” from documents.

# Identity Card



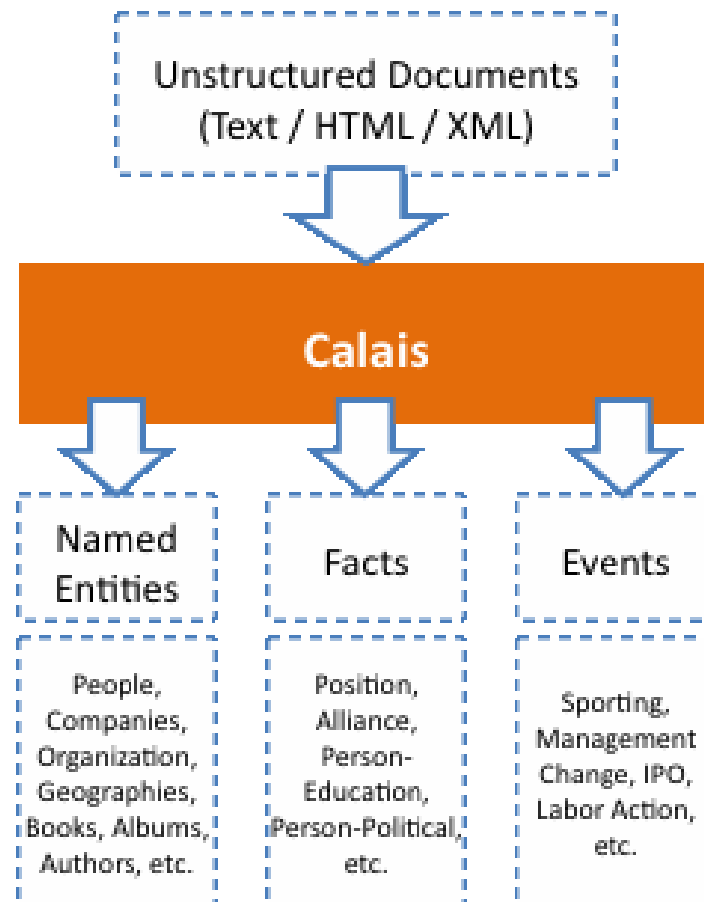
- Launched in January 2008.

- Powered by  THOMSON REUTERS

# How Does it Work?

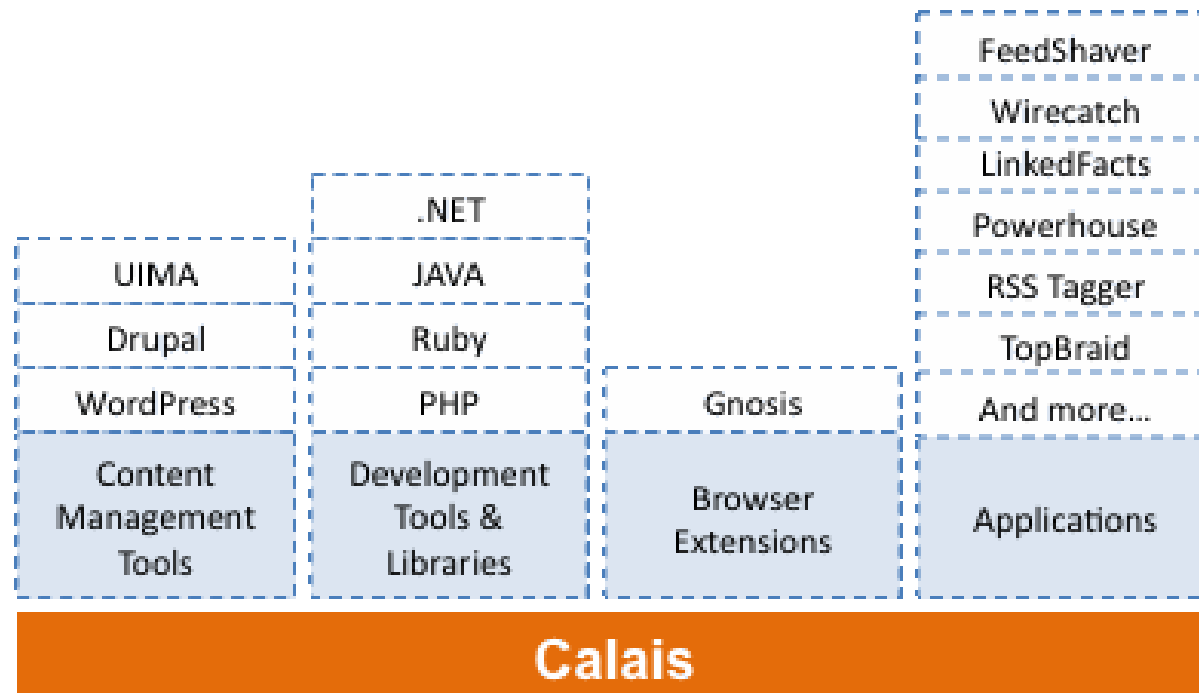
- It creates rich semantic metadata for the content you submit – in well under a second.
- Uses natural language processing, machine learning and other methods.
- Analyzes your document and finds the entities within it.

# How Does it Work?



# The Tools

- To deliver value to real people requires applications and the tools to build them.



# Live Cases

- Publishers and Bloggers

Alex Zettl and his group at the University of California, Berkeley report that they have developed an experimental memory device consisting of a crystalline iron nanoparticle that could have a storage capacity as high as 1 terabyte per square inch.

Calais

The screenshot displays the Calais analysis interface. It is divided into two main sections: 'Topics' and 'Entities'. The 'Topics' section shows 'Technology Internet: 100%' and 'Entertainment Culture: 69%'. The 'Entities' section lists four categories, each with a checked checkbox and a list of identified entities with their respective counts and relevance percentages.

Entity Type	Entity	Count	Relevance
Industry Term	experimental memory device	C:1	R:63%
Organization	University of California, Berkeley	C:1	R:63%
Person	Alex Zettl	C:2	R:69%
Province Or State	California, United States	C:1	R:63%



# Live Cases

- PowerHouse Museum



Online database with more than 60000 objects.

# OpenCalais on Drupal: What do you need?

- Calais module:  
<http://drupal.org/project/opencalais>.
- Resource Description Framework (RDF) module: <http://drupal.org/project/rdf>.
- ARC 2 RDF classes for PHP:  
<http://arc.semsol.org/download>.
- A Calais API key from <http://opencalais.com/>

# Insert the Key

Calais Configuration

Calais API Settings

Calais Node Settings

Calais Bulk Processing

## Calais API Key:

Required to utilize the Calais service. [Need One?](#)

## Calais Server:

The domain name for the server to use. Typically you will not have to change this unless you want to test beta functionality.

Allow Calais Searching

Indicates whether future searches can be performed on the extracted metadata by Calais

Allow Calais Distribution

Indicates whether the extracted metadata can be distributed by Calais

Save configuration

Reset to defaults

## Calais Configuration

[Calais API Settings](#)[Calais Node Settings](#)[Calais Bulk Processing](#)

### Store Calais RDF Locally

If checked, this will store the RDF data returned from Calais in the local RDF store. This data can be used to get a deeper view into the Calais results as the functionality is more fully developed. It also allows for future semantic applications to query and build upon this linked data.

—▷ [Global](#)

—▷ [Page](#)

▽ [Story](#)

#### Calais Processing:

- Not processed by Calais (default)
- Suggest terms, but DON'T apply them
- Apply all suggested terms on every update
- Apply all suggested terms on creation ONLY

#### Allow Calais Searching

Indicates whether future searches can be performed on the extracted metadata by Calais

#### Allow Calais Distribution

Indicates whether the extracted metadata can be distributed by Calais

#### Relevancy Threshold:

Determine how relevant a term must be in order for Calais to suggest it for a particular node. Based on a 0.00-1.00 scale, with 0.00 being least relevant (i.e. many terms appear).

#### Use Calais Global Entity defaults

If checked, this content type will use the Global Calais Entities.

## Calais Configuration

Calais API Settings

Calais Node Settings

Calais Bulk Processing

### Store Calais RDF Locally

If checked, this will store the RDF data returned from Calais in the local RDF store. This data can be used to get a deeper view into the Calais results as the functionality is more fully developed. It also allows for future semantic applications to query and build upon this linked data.

Global

### Which Calais Entities do you wish to use?:

- Anniversary
- CalaisDocumentCategory
- City
- Company
- Continent
- Country
- Currency
- EmailAddress
- EntertainmentAwardEvent
- EventsFacts
- Facility
- FaxNumber
- Holiday
- IndustryTerm
- MarketIndex
- MedicalCondition
- MedicalTreatment
- Movie

## Barcelona revels after Champions League win

[View](#)

[Edit](#)

[Calais](#)

Thu, 05/28/2009 - 14:54 — [admin](#)

Barcelona striker Thierry Henry said his side's 2-0 Champions League final triumph over Manchester United will stay long in the memory, while beaten goalkeeper Edwin van der Sar was at a loss to explain why his side failed to reproduce their top form.

Goals either side of half-time in the final at Rome's Olympic Stadium from Barcelona's Samuel Eto'o and Argentina star Lionel Messi gave the Spaniards their third European Cup crown.

Victory ensured Barcelona became the first team to win the Spanish treble of La Liga, the King's Cup and the Champions League.

Henry was in the Arsenal side which lost to Barcelona at Paris's Stade de France in 2006 and said the taste of Champions League victory was all the sweeter.

"This victory will remain engraved in my memory," said the 31-year-old.

"I had already known this kind of emotion when we won the Premier League title with Arsenal.

"It is something special to win the treble, it is a part of history."

Henry, who had double cause for celebration as it was his daughter's birthday, shrugged off a knee injury to play in Wednesday's final and said he left it until the eve of the game to decide whether he would play.

"For my part, I did not think I would be able to play, but on Tuesday evening I started to understand that it was going to be good," said the Frenchman.

"You just have to grit your teeth and get on with it, but as always with Barca, you always give your best.

**Country:**

Calais Suggestions: Argentina

**Events Facts:**

Calais Suggestions: Quotation Person Career

**Facility:**

Calais Suggestions: Olympic Stadium Stade de France

**Person:**

Calais Suggestions: **Thierry Henry** Edwin van der Sar Samuel Eto'o Lionel Messi La Liga Wayne Rooney Ryan Giggs Cristiano Ron.

**Position:**

Relevance: Moderate

Calais Suggestions: **goalkeeper** King Premier coach

**Sports Event:**

Calais Suggestions: **Champions League** Champions League final Olympic European Cup English Premier League

**Sports League:**

Calais Suggestions: **Champions League** Premier League English Premier League

**Anniversary:**

## Drupal Wins 2009 "Webware 100" Award for Social & Publishing

[View](#)[Edit](#)[Calais](#)

Award Recognizes Open Source Social Publishing Platform for its Leadership and Innovation

Berchem, Belgium (PRWEB) May 21, 2009 -- The Drupal Association today announced that Drupal has been named to the 2009 Webware 100, receiving top honors in the Social & Publishing category by Webware, a CNET site. With nearly 630,000 votes cast during the online voting poll this year, the Webware awards recognizes the best Web 2.0 sites, services, and applications on the web today. Drupal is once again a Webware 100 winner, receiving the award for a third consecutive year.

"As a community, we should be extremely proud of this award, and to be included amongst winners such as Facebook, Twitter, Wordpress, MySpace and Meebo is great validation for open source social publishing," said Dries Buytaert, creator and project lead of Drupal. "The Social & Publishing category is consistently one of the most competitive- this year garnering 185,000 votes- more votes than any other. Being ranked in the company of many of the most innovative Web 2.0 organizations and projects is once again a great moment for Drupal."

Drupal is an open source software project that makes it easy to publish content and build communities. Website administrators use Drupal to customize their website through a configurable user interface, the ability to extend the website with modules, and an extensible framework design. Drupal has been used on many popular and professional sites including Lifetime.com, FastCompany.com as well as, MTV's Web site in the United Kingdom, MTV.co.uk.

"The Webware 100 Award is a testament to amazing work being done by the Drupal community," added Tiffany Farriss, Special Projects Manager for the Drupal Association. "Being recognized for a third time demonstrates the strength and viability of Drupal as a leader in publishing technology, as well as the power of open source."

"This annual listing has become such an essential resource for people who want to learn more about Web 2.0 and the many ways it can help them stay connected," said Rafe Needleman, editor of Webware and CNET. "The Web 2.0 community and our editors ultimately selected the very best sites, services and applications that have made a lasting impression."

[Belgium](#) [CNET](#) [Dries Buytaert](#) [editor](#) [Facebook](#) [Innovation Berchem](#) [leader](#) [Manager for the Drupal Association](#) [MTV](#) [MTV.co.uk](#) [Person Career Quotation](#) [Rafe Needleman](#) [Special Projects](#) [Technology](#) [Tiffany Farriss](#) [Twitter](#) [United Kingdom](#) [Webware](#)