ENABLING SEMANTIC APPLICATIONS THROUGH CALAIS

Summer School on Multimedia Semantics
September 2008
Crete, Greece
A QUICK 45 MINUTES

INTRODUCING THOMSON REUTERS

- WHAT WE MEAN BY ‘INTELLIGENT INFORMATION’
- A STRATEGY FOR CONTENT INTERPERABILITY

INTRODUCING OPEN CALAIS

- TOP CALAIS APPLICATIONS
- FIVE EASY WAYS TO GET STARTED
THOMSON REUTERS: TARGETING EVERY PROFESSIONAL ON THE PLANET

<table>
<thead>
<tr>
<th>Region</th>
<th>Population</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Americas</td>
<td>27,621</td>
<td>54%</td>
</tr>
<tr>
<td>Europe</td>
<td>12,018</td>
<td>23%</td>
</tr>
<tr>
<td>Asia</td>
<td>12,322</td>
<td>23%</td>
</tr>
</tbody>
</table>
THOMSON REUTERS AT A GLANCE

<table>
<thead>
<tr>
<th>Financial</th>
<th>Provides financial applications for over half a million professionals globally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legal</td>
<td>Westlaw relied upon by 98% of the world’s major law firms</td>
</tr>
<tr>
<td>Tax &amp; Accounting</td>
<td>Checkpoint used by 99 of the top 100 U.S. accounting firms</td>
</tr>
<tr>
<td>Scientific</td>
<td>Used by over 20 million researchers worldwide</td>
</tr>
<tr>
<td>Healthcare</td>
<td>The largest U.S. hospital decision support provider</td>
</tr>
<tr>
<td>Media</td>
<td>Reuters News reaches over one billion people daily</td>
</tr>
</tbody>
</table>
WHAT IS INTELLIGENT INFORMATION?

• Perfect support of the underlying task

• The II for a particular situation probably must be created in real time
## The Content Landscape

<table>
<thead>
<tr>
<th></th>
<th>TR Products</th>
<th>Non-TR Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>TR Content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Licensed (3rd party) content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-licensed content (Web)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customers’ proprietary content</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TR’S WALLED GARDEN OF DATA
INTRODUCING OPEN CALAIS: EVERYONE WINS

Powerful, free markup services to anyone who creates, aggregates or consumes content.

Standard, simple access to unwashed Web content for all of TR’s products and services.

End users get closer to information they need, vs. data that’s available.
OPEN CALAIS OVERVIEW

The Process

Calais accepts unstructured text and uses sophisticated NLP and machine learning techniques to return intelligent metadata.
The Toolkit

Isn’t just a web service – but a growing ecosystem of libraries, plugins and applications designed to ease adoption

OPEN CALAIS OVERVIEW
OPEN CLAIS – NINE MONTHS IN…

- Launched in late January 2008
- Already, 5,700 developers have joined
- Approx. 500,000 content ‘transactions’ per day
- Delivered two major update releases
- Lots of interesting apps
  - The Mail & Guardian Online (http://www.mg.co.za/)
  - www.powerhousemuseum.com
  - Gist.whistlehog.com
  - www.opencalais.com/gallery
FIVE EASY WAYS TO GET STARTED

• Use Calais to automatically tag your archives
  = improved search & navigation

• Use Calais to create microformat metadata for Yahoo! SearchMonkey
  = better search and usability

• Use Calais to drive alerts or feeds based on events, not just keywords
  = timely alerts on events your users care about

• Use Calais to enable semantic knowledge discovery
  = derive valuable content insights

• Use Calais’ toolkit of end-user ready applications
  = no coding required and you see immediate benefits
Questions?

A quick hit list of what’s ahead for Calais:

• Continued rapid release of new tools & applications
• Major content partnerships
• Calais: Connected
• Calais Metadata Schema made public
• Four additional languages
Thank You!

Barak Pridor, CEO, ClearForest, a Thomson Reuters Company
Example: The Mail & Guardian Online, South Africa

Using Calais to optimize search and navigation; drive consumer engagement

South Africa’s Mail & Guardian used Calais to metatag both new and historical articles. This enabled them to:

1. Build an index or topics A-Z
2. Pull out automatic related articles or pictures
3. Create news alerts on companies or people
4. Pull up maps for the countries named in articles
5. Predict readers’ interests based on browsing habits
6. Create tag clouds, showing popular subjects, people, etc.
Example: Gist - today’s news filtered by people, places & events

Using Calais to prioritize stories, rank newsmakers & reveal trends / reader demand
Example: The Powerhouse Museum in Sydney

Using Calais to tag historical archives & using tags as search terms

![Image of theodolite]

**Statement of significance**

When George Emanuel Davison died in 1963, at age 94, he was the oldest Registered Surveyor in New South Wales. He was born in 1869 at Armidale. He joined the NSW Lands Department in 1906 as a cadet surveyor and then served articles with Surveyors Alcock and Hannett and qualified on 14 April 1914 (Certificate 1222). He rejoined the Lands Department and worked with them until March 1950. He then continued to work as a freelance surveyor doing his last work in Kangaroo Valley on behalf of the Sydney Bush Walking Club when he was 68 years old.

He purchased this theodolite some time in the late 1920s and he used it for the rest of his surveying career. The theodolite is not remarkable in its own right being a standard instrument of the time but it is part of a collection of surveying equipment and an archive of papers and photographs that illustrate the working life of a surveyor in the first half of the 19th century.

During his time in the field with the Lands Department George Davison usually lived under canvas and travelled by horseback. He would be away from home for months and sometimes his wife would accompany him. There are photographs showing the camp and George and his wife's tent including her personal effects.

George worked in many areas of New South Wales including the areas around Hay and Wagga Wagga, The Snowy Mountains, Armidale, Grafton, Forbes and Condobolin. His topographic work done in the Snowy Mountains was commenced at the opening of the Snowy Mountains Hydro-Electric Scheme. SMHEA engineers were very impressed at the accuracy of the maps produced.