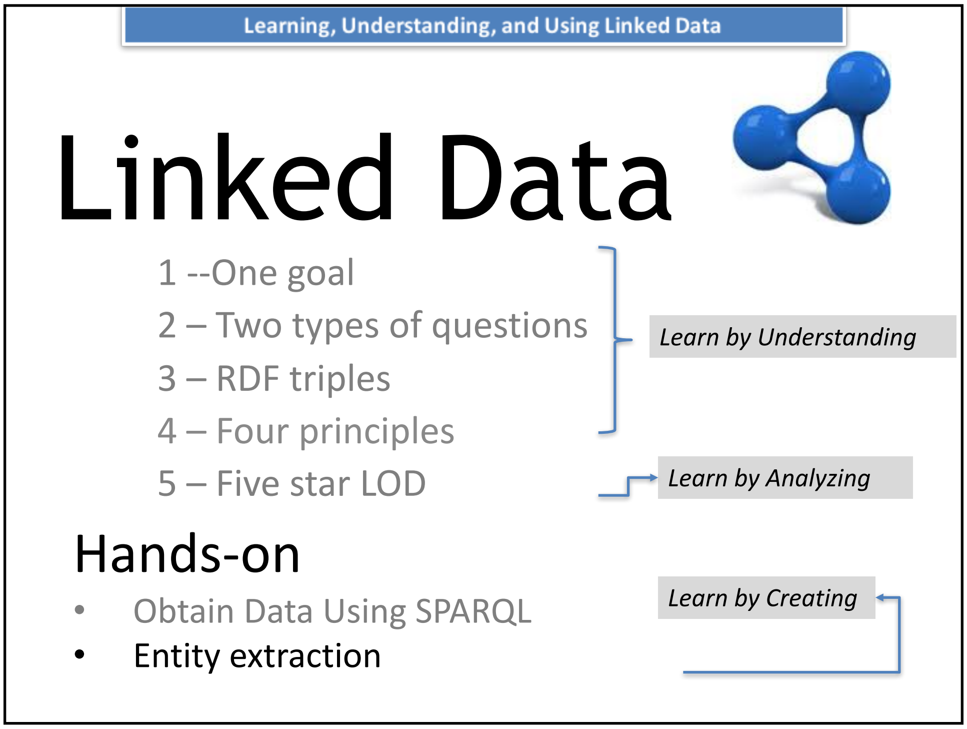


Learning, Understanding, and Using Linked Data

Linked Data

1 --One goal
2 – Two types of questions
3 – RDF triples
4 – Four principles
5 – Five star LOD

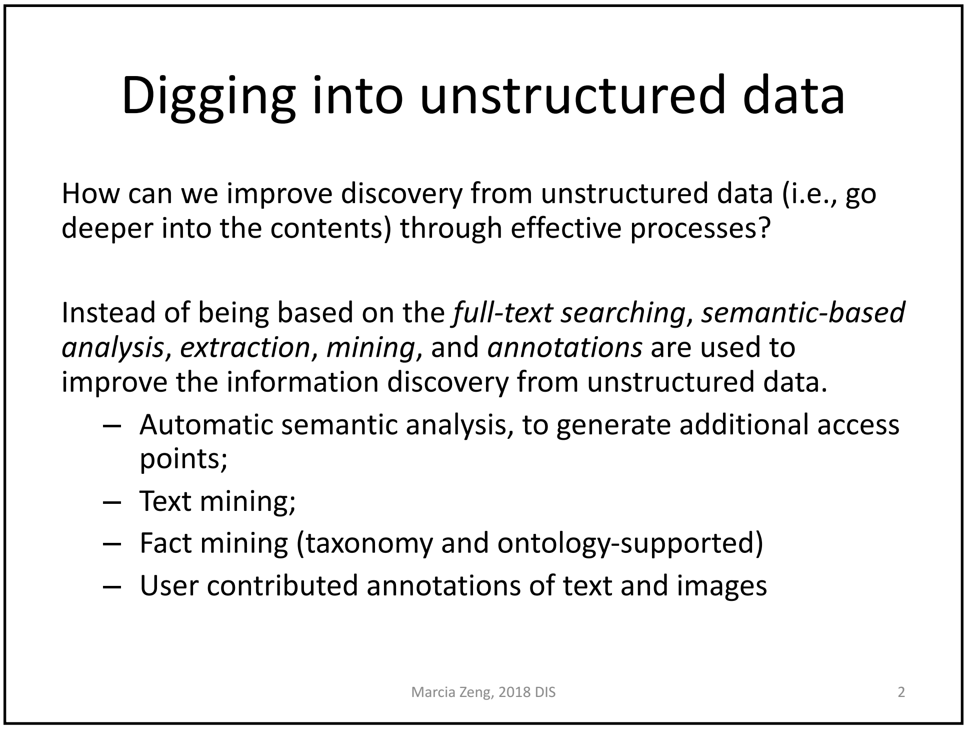
Learn by Understanding

Learn by Analyzing

Hands-on

- Obtain Data Using SPARQL
- Entity extraction

Learn by Creating



Digging into unstructured data

How can we improve discovery from unstructured data (i.e., go deeper into the contents) through effective processes?

Instead of being based on the *full-text searching*, *semantic-based analysis*, *extraction*, *mining*, and *annotations* are used to improve the information discovery from unstructured data.

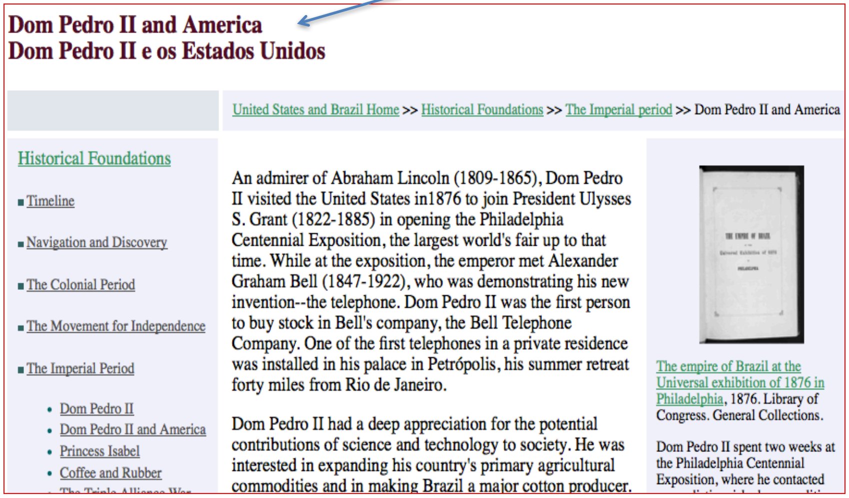
– Automatic semantic analysis, to generate additional access points;
– Text mining;
– Fact mining (taxonomy and ontology-supported)
– User contributed annotations of text and images

An example

Many archival materials are digitized. But they are still unstructured, just like this interesting website:

http://international.loc.gov/intldl/brhtml/br-1/br-1-5-2.html#track1

Now the only name that is obvious is Dom Pedro II.



Dom Pedro II and America
Dom Pedro II e os Estados Unidos

United States and Brazil Home >> Historical Foundations >> The Imperial period >> Dom Pedro II and America

Historical Foundations

- Timeline
- Navigation and Discovery
- The Colonial Period
- The Movement for Independence
- The Imperial Period
  - Dom Pedro II
  - Dom Pedro II and America
  - Princess Isabel
  - Coffee and Rubber
  - The Triple Alliance War

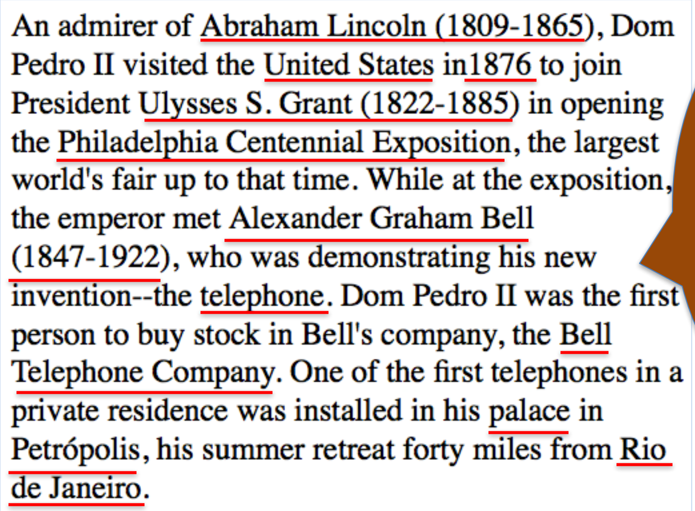
An admirer of Abraham Lincoln (1809-1865), Dom Pedro II visited the United States in1876 to join President Ulysses S. Grant (1822-1885) in opening the Philadelphia Centennial Exposition, the largest world's fair up to that time. While at the exposition, the emperor met Alexander Graham Bell (1847-1922), who was demonstrating his new invention--the telephone. Dom Pedro II was the first person to buy stock in Bell's company, the Bell Telephone Company. One of the first telephones in a private residence was installed in his palace in Petrópolis, his summer retreat forty miles from Rio de Janeiro.

Dom Pedro II had a deep appreciation for the potential contributions of science and technology to society. He was interested in expanding his country's primary agricultural commodities and in making Brazil a major cotton producer.

The empire of Brazil at the Universal exhibition of 1876 in Philadelphia, 1876. Library of Congress. General Collections.

Dom Pedro II spent two weeks at the Philadelphia Centennial Exposition, where he contacted

Think: In this document, how many interesting things people would like to know more about…



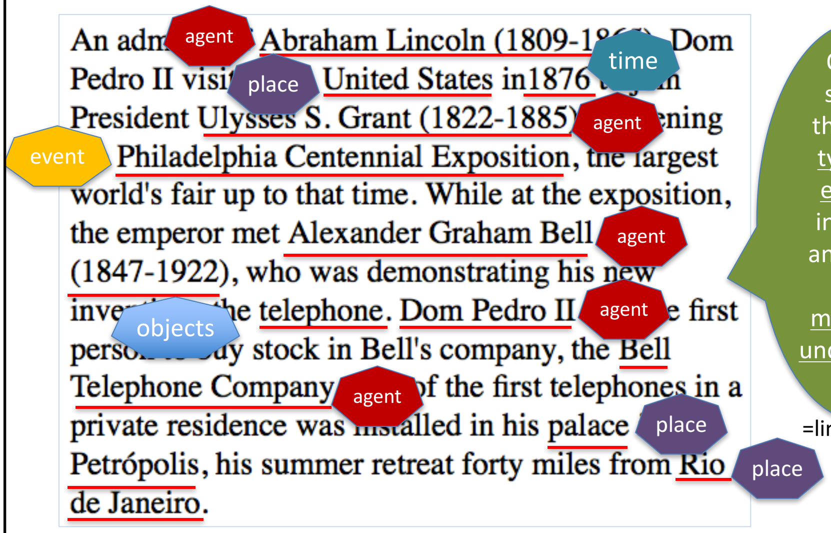
An admirer of Abraham Lincoln (1809-1865), Dom Pedro II visited the United States in1876 to join President Ulysses S. Grant (1822-1885) in opening the Philadelphia Centennial Exposition, the largest world's fair up to that time. While at the exposition, the emperor met Alexander Graham Bell (1847-1922), who was demonstrating his new invention--the telephone. Dom Pedro II was the first person to buy stock in Bell's company, the Bell Telephone Company. One of the first telephones in a private residence was installed in his palace in Petrópolis, his summer retreat forty miles from Rio de Janeiro.

We can [manually]make hyperlinks under strings … (machine-processable)

= linking strings

Annotation

From linking strings →to→ linking things


agent
place
time
agent
event
agent
objects
agent
agent
place
place

An adm Abraham Lincoln (1809-18), Dom
Pedro II visi United States in1876 t n
President Ulysses S. Grant (1822-1885) ening
Philadelphia Centennial Exposition, the largest
world's fair up to that time. While at the exposition,
the emperor met Alexander Graham Bell
(1847-1922), who was demonstrating his new
inve he telephone. Dom Pedro II e first
pers uy stock in Bell's company, the Bell
Telephone Company of the first telephones in a
private residence was lled in his palace
Petrópolis, his summer retreat forty miles from Rio
de Janeiro.

Or, we should think the types of entities involved and make them machine-understandable

=linking things

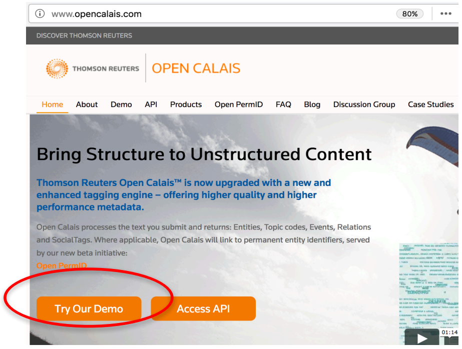
THOMSON REUTERS | OPEN CALAIS

Semantic Analysis Tool #1

Open Calais

"Open Calais processes the text you submit and returns: Entities, Topic codes, Events, Relations and SocialTags. Where applicable, Open Calais will link to permanent entity identifiers, served by our new beta initiative:"

http://www.opencalais.com/



Bring Structure to Unstructured Content

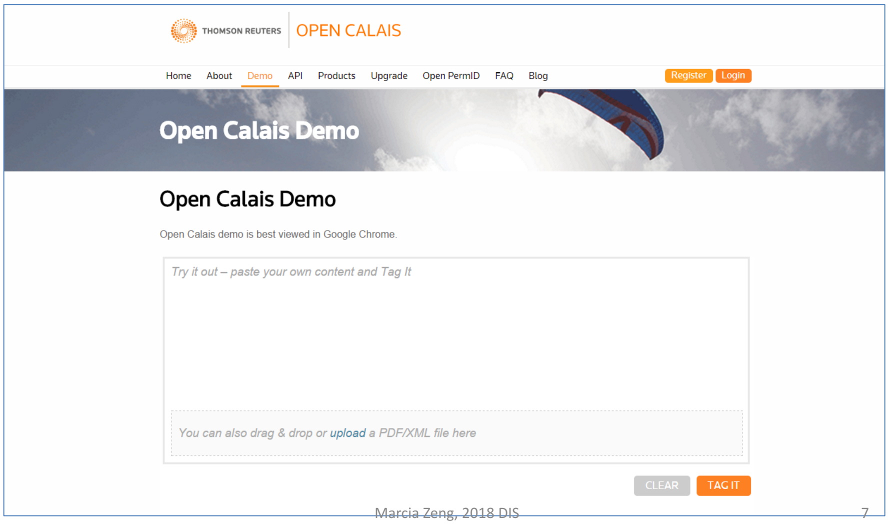
Thomson Reuters Open Calais™ is now upgraded with a new and enhanced tagging engine – offering higher quality and higher performance metadata.

Open Calais processes the text you submit and returns: Entities, Topic codes, Events, Relations and SocialTags. Where applicable, Open Calais will link to permanent entity identifiers, served by our new beta initiative:

THOMSON REUTERS | OPEN CALAIS

Demo: Using Open Calais (free version) tool to run through the text of that paragraph showing previously.

1. Go to: http://www.opencalais.com/opencalais-demo/



Open Calais Demo

Open Calais demo is best viewed in Google Chrome.

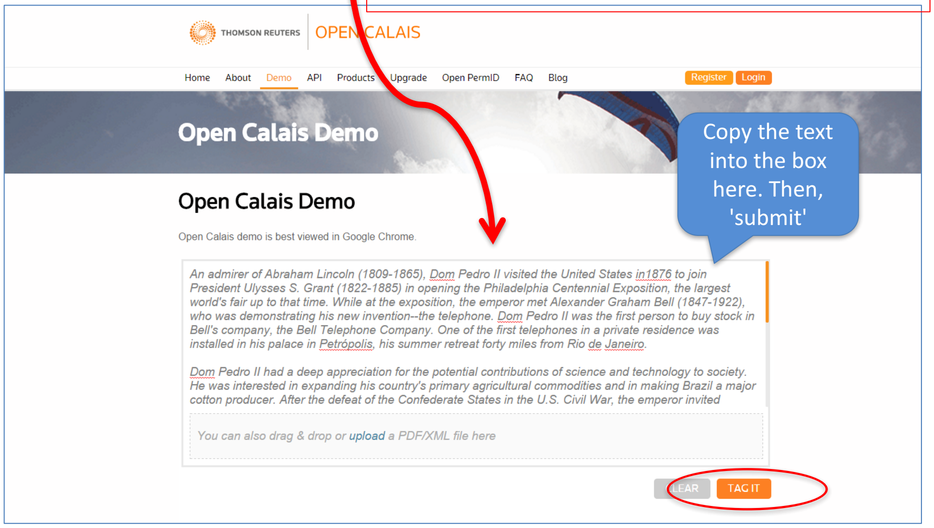
Try it out – paste your own content and Tag It

You can also drag & drop or upload a PDF/XML file here

2. Copy the text into the box. "Submit"

I used the same text used in my example. If you use another text, do not use private or unclassified material.

THOMSON REUTERS | OPEN CALAIS



Copy the text into the box here. Then, 'submit'

Open Calais Demo

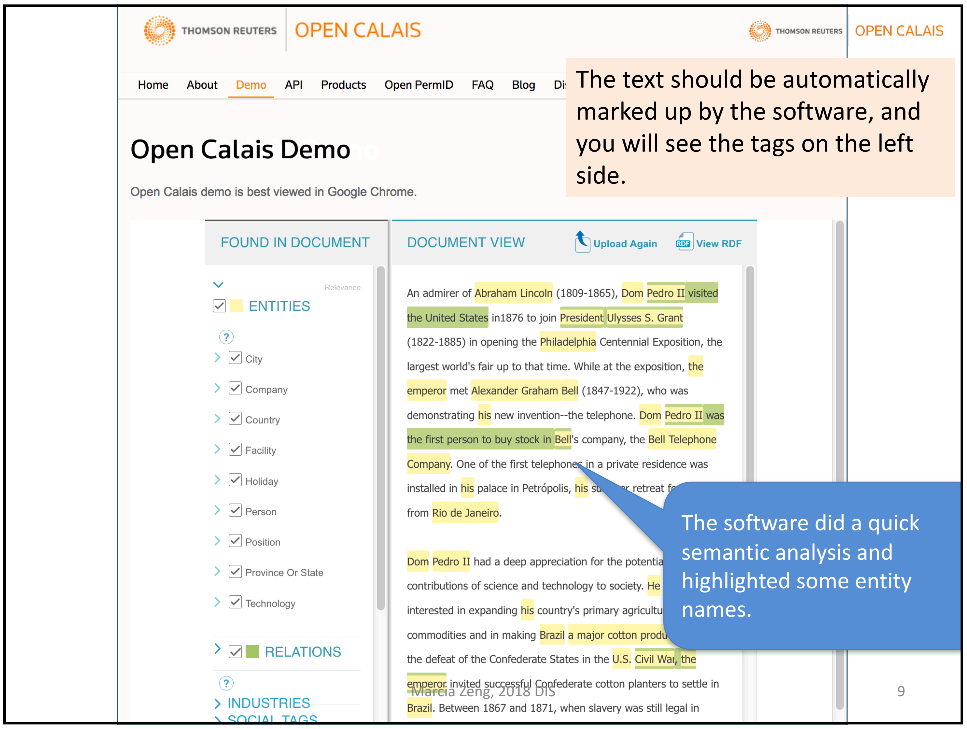
Open Calais demo is best viewed in Google Chrome.

An admirer of Abraham Lincoln (1809-1865), Dom Pedro II visited the United States in1876 to join President Ulysses S. Grant (1822-1885) in opening the Philadelphia Centennial Exposition, the largest world's fair up to that time. While at the exposition, the emperor met Alexander Graham Bell (1847-1922), who was demonstrating his new invention--the telephone. Dom Pedro II was the first person to buy stock in Bell's company, the Bell Telephone Company. One of the first telephones in a private residence was installed in his palace in Petrópolis, his summer retreat forty miles from Rio de Janeiro.

Dom Pedro II had a deep appreciation for the potential contributions of science and technology to society. He was interested in expanding his country's primary agricultural commodities and in making Brazil a major cotton producer. After the defeat of the Confederate States in the U.S. Civil War, the emperor invited

You can also drag & drop or upload a PDF/XML file here

http://www.opencalais.com/opencalais-demo/



The text should be automatically marked up by the software, and you will see the tags on the left side.

Open Calais Demo

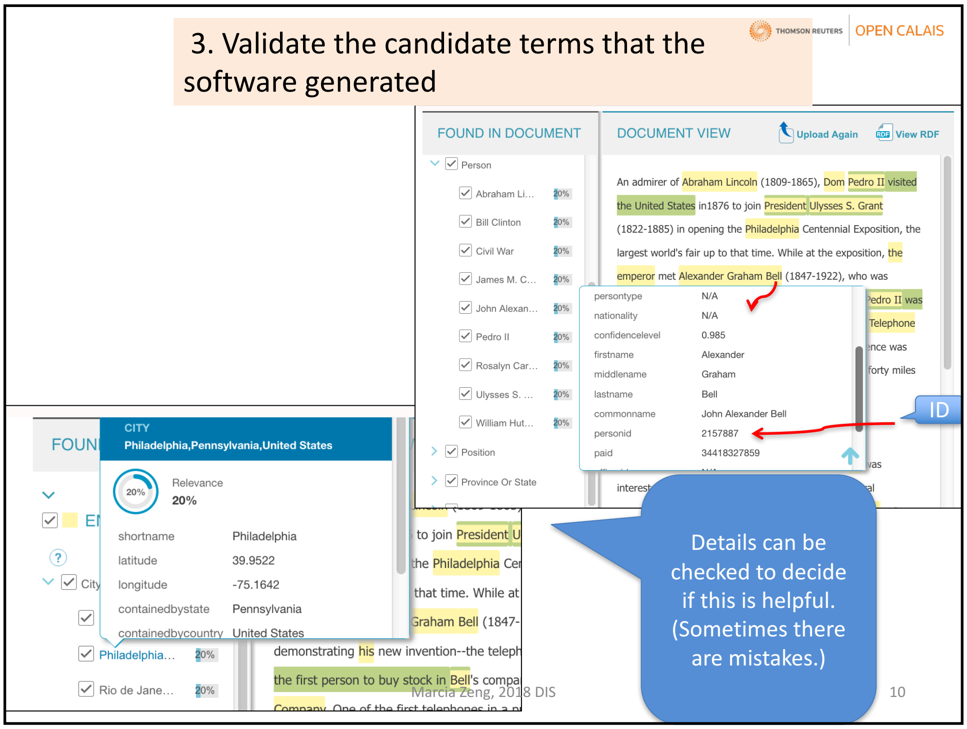
Open Calais demo is best viewed in Google Chrome.

An admirer of Abraham Lincoln (1809-1865), Dom Pedro II visited the United States in1876 to join President Ulysses S. Grant (1822-1885) in opening the Philadelphia Centennial Exposition, the largest world's fair up to that time. While at the exposition, the emperor met Alexander Graham Bell (1847-1922), who was demonstrating his new invention--the telephone. Dom Pedro II was the first person to buy stock in Bell's company, the Bell Telephone Company. One of the first telephones in a private residence was installed in his palace in Petrópolis, his su er retreat f from Rio de Janeiro.

Dom Pedro II had a deep appreciation for the potentia contributions of science and technology to society. He interested in expanding his country's primary agricultu commodities and in making Brazil a major cotton produ the defeat of the Confederate States in the U.S. Civil War, the emperor invited successful Confederate cotton planters to settle in Brazil. Between 1867 and 1871, when slavery was still legal in

The software did a quick semantic analysis and highlighted some entity names.

Marcia Zeng, 2018 DIS 9



3. Validate the candidate terms that the software generated

An admirer of Abraham Lincoln (1809-1865), Dom Pedro II visited the United States in1876 to join President Ulysses S. Grant (1822-1885) in opening the Philadelphia Centennial Exposition, the largest world's fair up to that time. While at the exposition, the emperor met Alexander Graham Bell (1847-1922), who was

persontype	N/A
nationality	N/A
confidencelevel	0.985
firstname	Alexander
middlename	Graham
lastname	Bell
commonname	John Alexander Bell
personid	2157887
paid	34418327859

ID

CITY
Philadelphia,Pennsylvania,United States

Relevance
20%

shortname	Philadelphia
latitude	39.9522
longitude	-75.1642
containedbystate	Pennsylvania
containedbycountry	United States

to join President U the Philadelphia Ce that time. While at Graham Bell (1847- demonstrating his new invention--the teleph the first person to buy stock in Bell's compa Company. One of the first telephones in a p

Details can be checked to decide if this is helpful. (Sometimes there are mistakes.)

Marcia Zeng, 2018 DIS 10

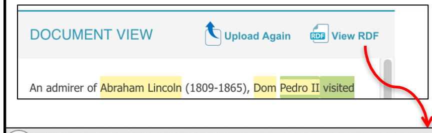
THOMSON REUTERS | OPEN CALAIS

Behind the theme

Let's check out what this looks like at the back-end

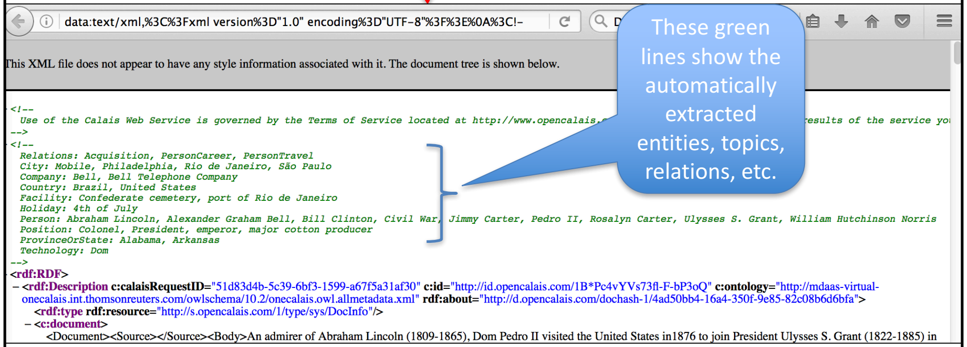
What are the structured data behind these?
(Coded with RDF/XML syntax. [You do not need to understand the codes at the moment.])

THOMSON REUTERS | OPEN CALAIS



An admirer of Abraham Lincoln (1809-1865), Dom Pedro II visited

Clicking on "View RDF" brings us to the machine-processable datasets.


data:text/xml,%3C%3Fxml version%3D"1.0" encoding%3D"UTF-8"%3F%3E%0A%3C!-

This XML file does not appear to have any style information associated with it. The document tree is shown below.

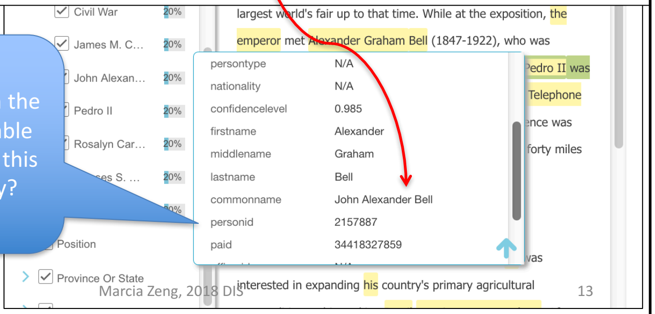
```
<!--
  Use of the Calais Web Service is governed by the Terms of Service located at http://www.opencalais.c           results of the service yo
-->
<!--
  Relations: Acquisition, PersonCareer, PersonTravel
  City: Mobile, Philadelphia, Rio de Janeiro, São Paulo
  Company: Bell, Bell Telephone Company
  Country: Brazil, United States
  Facility: Confederate cemetery, port of Rio de Janeiro
  Holiday: 4th of July
  Person: Abraham Lincoln, Alexander Graham Bell, Bill Clinton, Civil War, Jimmy Carter, Pedro II, Rosalyn Carter, Ulysses S. Grant, William Hutchinson Norris
  Position: Colonel, President, emperor, major cotton producer
  ProvinceOrState: Alabama, Arkansas
  Technology: Dom
-->
<rdf:RDF>
 – <rdf:Description c:calaisRequestID="51d83d4b-5c39-6bf3-1599-a67f5a31af30" c:id="http://id.opencalais.com/1B*Pc4vYVs73fl-F-bP3oQ" c:ontology="http://mdaas-virtual-
   onecalais.int.thomsonreuters.com/owlschema/10.2/onecalais.owl.allmetadata.xml" rdf:about="http://d.opencalais.com/dochash-1/4ad50bb4-16a4-350f-9e85-82c08b6d6bfa">
     <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/DocInfo"/>
   – <c:document>
       <Document><Source></Source><Body>An admirer of Abraham Lincoln (1809-1865), Dom Pedro II visited the United States in1876 to join President Ulysses S. Grant (1822-1885) in
```

These green lines show the automatically extracted entities, topics, relations, etc.

THOMSON REUTERS | OPEN CALAIS

```
– <rdf:Description rdf:about="http://d.opencalais.com/er/person/ralg-pa1/3441832
    <rdf:type rdf:resource="http://s.opencalais.com/1/type/er/Person"/>
    <c:docId rdf:resource="http://d.opencalais.com/dochash-1/4ad50bb4-16a4-35(
    <c:name>John Alexander Bell</c:name>
    <c:personid>2157887</c:personid>
    <c:paid>34418327859</c:paid>
    <c:officerid>N/A</c:officerid>
    <c:commonname>John Alexander Bell</c:commonname>
    <c:score>0.6591502</c:score>
    <!--Alexander Graham Bell-->
    <c:subject rdf:resource="http://d.opencalais.com/pershash-1/39ff7fbd-7150-3225-be2d-a15d6fce1d34"/>
  </rdf:Description>
```

If you search "Bell" on this code page, you can find several description sets, including this one, coded his name, officer–ID, etc.



☑ Civil War 20%
☑ James M. C… 20%
☑ John Alexan… 20%
☑ Pedro II 20%
☑ Rosalyn Car… 20%
…es S. … 20%
Position
☑ Province Or State

persontype	N/A
nationality	N/A
confidencelevel	0.985
firstname	Alexander
middlename	Graham
lastname	Bell
commonname	John Alexander Bell
personid	2157887
paid	34418327859

largest world's fair up to that time. While at the exposition, the emperor met Alexander Graham Bell (1847-1922), who was … Pedro II was … Telephone … ence was … forty miles … was … interested in expanding his country's primary agricultural

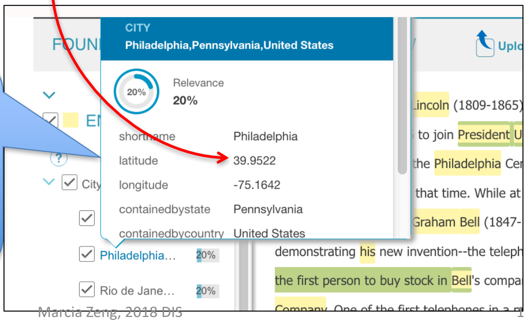
Remember on the human-readable page, we saw this data already?

Marcia Zeng, 2018 DIS 13

THOMSON REUTERS | OPEN CALAIS

```
– <rdf:Description rdf:about="http://d.opencalais.com/er/geo/city/ralg-geo1/06db416f-aa27-f883-8bbc-e038dbfbde9d">
    <rdf:type rdf:resource="http://s.opencalais.com/1/type/er/Geo/City"/>
    <c:docId rdf:resource="http://d.opencalais.com/dochash-1/4ad50bb4-16a4-350f-9e85-82c08b6d6bfa"/>
    <c:name>Philadelphia,Pennsylvania,United States</c:name>
    <c:shortname>Philadelphia</c:shortname>
    <c:latitude>39.9522</c:latitude>
    <c:longitude>-75.1642</c:longitude>
    <c:containedbystate>Pennsylvania</c:containedbystate>
    <c:containedbycountry>United States</c:containedbycountry>
    <!--Philadelphia-->
    <c:subject rdf:resource="http://d.opencalais.com/genericHasher-1/4b474cf2-e8df-3193-86a4-8e858a5c5e37"/>
    <c:rcscode>G:22H</c:rcscode>
  </rdf:Description>
```

Same for places, e.g., "Philadelphia". Can you identify its latitude and longitude?



FOUN…

CITY
Philadelphia,Pennsylvania,United States

Relevance
20%

shortname	Philadelphia
latitude	39.9522
longitude	-75.1642
containedbystate	Pennsylvania
containedbycountry	United States

☑ City
☑ Philadelphia… 20%
☑ Rio de Jane… 20%

Lincoln (1809-1865) … to join President U… the Philadelphia Ce… that time. While at … Graham Bell (1847- … demonstrating his new invention--the teleph… the first person to buy stock in Bell's compa… Company. One of the first telephones in a …

Again, remember in the human-readable page, we saw this data already?

Marcia Zeng, 2018 DIS 14

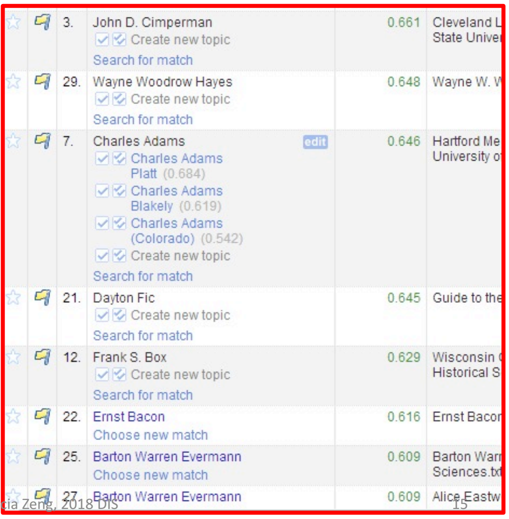
How can we use this data? (Technical processes).

1. If you know programming, you can convert this data into a table.
2. The RDF/XML file can be converted into a different format, and stored.

OPEN Refine

3. For individual or multiple sets of data, use OpenRefine http://openrefine.org/ to clean up and reconcile.

E.g., merge synonyms of an entity, such as, "Kent State", "KSU", "Kent State University", "Kent State U.", etc.



3.	John D. Cimperman ☑ ☑ Create new topic Search for match	0.661	Cleveland L State Unive
29.	Wayne Woodrow Hayes ☑ ☑ Create new topic Search for match	0.648	Wayne W. W
7.	Charles Adams edit ☑ ☑ Charles Adams Platt (0.684) ☑ ☑ Charles Adams Blakely (0.619) ☑ ☑ Charles Adams (Colorado) (0.542) ☑ ☑ Create new topic Search for match	0.646	Hartford Me University o
21.	Dayton Fic ☑ ☑ Create new topic Search for match	0.645	Guide to the
12.	Frank S. Box ☑ ☑ Create new topic Search for match	0.629	Wisconsin (Historical S
22.	Ernst Bacon Choose new match	0.616	Ernst Bacor
25.	Barton Warren Evermann Choose new match	0.609	Barton Warr Sciences.txt
27.	Barton Warren Evermann	0.609	Alice Eastw

After merging, the named entity can go through the reconcile process; OpenRefine also enables you to obtain URIs from name authorities and thesauri through API.

Let me know if you want an example from my research of using OpenRefine.

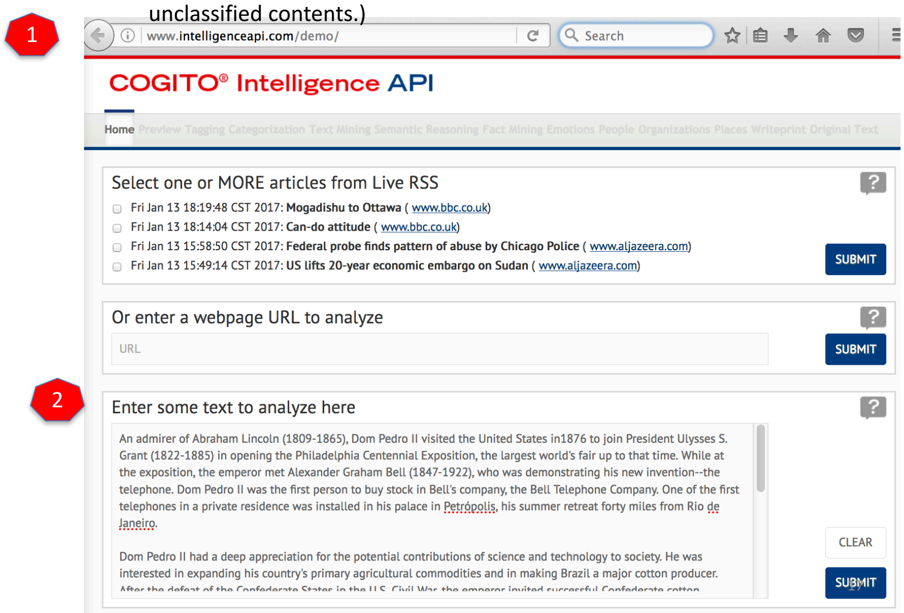
Additional Information

- Cogito Intelligence http://www.intelligenceapi.com/demo/ for for Intelligence, Terrorism, Cyber Crime, Crime, Geographic domains, etc.
- NIH's MetaMap: https://metamap.nlm.nih.gov/ for biomedical and health related domains.

COGITO® Intelligence API

Anther tool:

1. Go to http://www.intelligenceapi.com/demo/
2. Paste the same text from my example.
   (You may use your own text from other places. Do not use private or unclassified contents.)


1
2
www.intelligenceapi.com/demo/
Search
COGITO® Intelligence API
Home Preview Tagging Categorization Text Mining Semantic Reasoning Fact Mining Emotions People Organizations Places Writeprint Original Text
?
SUBMIT
URL
CLEAR

Select one or MORE articles from Live RSS

- □ Fri Jan 13 18:19:48 CST 2017: **Mogadishu to Ottawa** (www.bbc.co.uk)
- □ Fri Jan 13 18:14:04 CST 2017: **Can-do attitude** (www.bbc.co.uk)
- □ Fri Jan 13 15:58:50 CST 2017: **Federal probe finds pattern of abuse by Chicago Police** (www.aljazeera.com)
- □ Fri Jan 13 15:49:14 CST 2017: **US lifts 20-year economic embargo on Sudan** (www.aljazeera.com)

Or enter a webpage URL to analyze

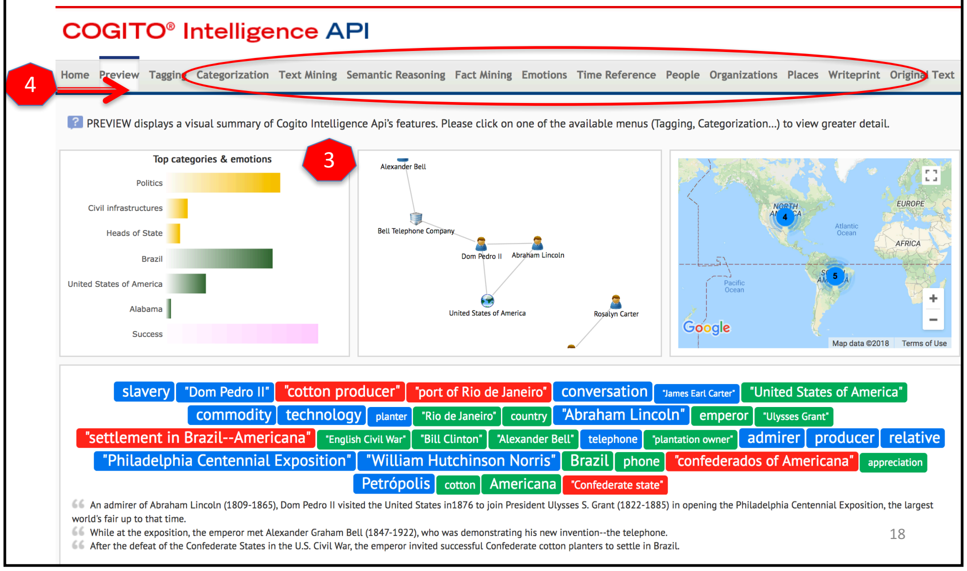
Enter some text to analyze here

An admirer of Abraham Lincoln (1809-1865), Dom Pedro II visited the United States in1876 to join President Ulysses S. Grant (1822-1885) in opening the Philadelphia Centennial Exposition, the largest world's fair up to that time. While at the exposition, the emperor met Alexander Graham Bell (1847-1922), who was demonstrating his new invention--the telephone. Dom Pedro II was the first person to buy stock in Bell's company, the Bell Telephone Company. One of the first telephones in a private residence was installed in his palace in Petrópolis, his summer retreat forty miles from Rio de Janeiro.

Dom Pedro II had a deep appreciation for the potential contributions of science and technology to society. He was interested in expanding his country's primary agricultural commodities and in making Brazil a major cotton producer.
After the defeat of the Confederate States in the U.S. Civil War, the emperor invited successful Confederate cotton

COGITO® Intelligence API

3. See what the machine did on the preview page.
4. Click on these one-by-one and think of how many FACETS the tool has used.



PREVIEW displays a visual summary of Cogito Intelligence Api's features. Please click on one of the available menus (Tagging, Categorization...) to view greater detail.

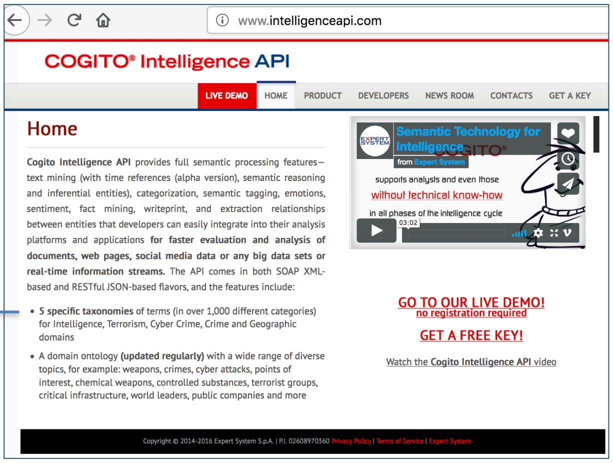
> An admirer of Abraham Lincoln (1809-1865), Dom Pedro II visited the United States in1876 to join President Ulysses S. Grant (1822-1885) in opening the Philadelphia Centennial Exposition, the largest world's fair up to that time.
>
> While at the exposition, the emperor met Alexander Graham Bell (1847-1922), who was demonstrating his new invention--the telephone.
>
> After the defeat of the Confederate States in the U.S. Civil War, the emperor invited successful Confederate cotton planters to settle in Brazil.

18

COGITO® Intelligence API

Behind the theme

- **Five (5) specific taxonomies** of terms (in over 1,000 different categories):
  - Intelligence,
  - Terrorism,
  - Cyber Crime,
  - Crime, and
  - Geographic domains
- **A domain ontology** (updated regularly) with a wide range of diverse topics



Home

Cogito Intelligence API provides full semantic processing features—text mining (with time references (alpha version), semantic reasoning and inferential entities), categorization, semantic tagging, emotions, sentiment, fact mining, writeprint, and extraction relationships between entities that developers can easily integrate into their analysis platforms and applications **for faster evaluation and analysis of documents, web pages, social media data or any big data sets or real-time information streams.** The API comes in both SOAP XML-based and RESTful JSON-based flavors, and the features include:

- **5 specific taxonomies** of terms (in over 1,000 different categories) for Intelligence, Terrorism, Cyber Crime, Crime and Geographic domains
- A domain ontology **(updated regularly)** with a wide range of diverse topics, for example: weapons, crimes, cyber attacks, points of interest, chemical weapons, controlled substances, terrorist groups, critical infrastructure, world leaders, public companies and more

GO TO OUR LIVE DEMO!
no registration required

GET A FREE KEY!

Watch the **Cogito Intelligence API** video

Copyright © 2014-2016 Expert System S.p.A. | P.I. 02608970360 Privacy Policy | Terms of Service | Expert System

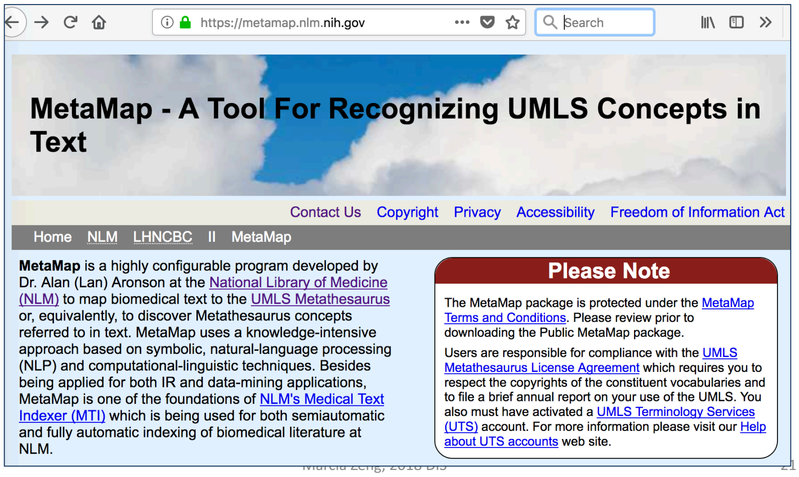
How can I use these tools in the future?

- For each of the tools, you can apply for a free “key” so that you can conduct analysis of longer text and patch process for folders and files.
- These tools might not be the best for humanities domains. But they (machine) learn quickly.
- Any text submitted will be kept in their corpus, so, **do not** submit any private or unclassified documents!

Anther tool:

NIH’s MetaMap: https://metamap.nlm.nih.gov/ for biomedical and health related domains.

Supported by *Metathesaurus*, a large biomedical thesaurus that is organized by concept or meaning, and it links similar names for the same concept from nearly 200 different vocabularies from the world.



MetaMap - A Tool For Recognizing UMLS Concepts in Text

Contact Us Copyright Privacy Accessibility Freedom of Information Act

Home NLM LHNCBC II MetaMap

MetaMap is a highly configurable program developed by Dr. Alan (Lan) Aronson at the National Library of Medicine (NLM) to map biomedical text to the UMLS Metathesaurus or, equivalently, to discover Metathesaurus concepts referred to in text. MetaMap uses a knowledge-intensive approach based on symbolic, natural-language processing (NLP) and computational-linguistic techniques. Besides being applied for both IR and data-mining applications, MetaMap is one of the foundations of NLM's Medical Text Indexer (MTI) which is being used for both semiautomatic and fully automatic indexing of biomedical literature at NLM.

Please Note

The MetaMap package is protected under the MetaMap Terms and Conditions. Please review prior to downloading the Public MetaMap package.

Users are responsible for compliance with the UMLS Metathesaurus License Agreement which requires you to respect the copyrights of the constituent vocabularies and to file a brief annual report on your use of the UMLS. You also must have activated a UMLS Terminology Services (UTS) account. For more information please visit our Help about UTS accounts web site.