COGITO Intelligence API: Main Features

Release 9.0, March 2015

CATEGORIZATION

COGITO Intelligence API categorizes texts according to different taxonomies. The standard version offers three different **Automatic Categorization** engines.

- **Intelligence** categorization engine: Developed to support the activities of Intelligence Agencies that are required to quickly and accurately assess large amounts of diverse information.
- **Crime & Offense** categorization: Taxonomy focused on the needs of Law Enforcement Agencies.
- **Geotaxonomy** categorization: Categorizes content for geographic locations and information.
- **Computer Crime** categorization: Detailed taxonomy designed for accurate categorization of the Cyber Illegal domain.

With these three taxonomies, **COGITO Intelligence API** uses a comprehensive approach to classify the contents of more than 1000 different items, organized hierarchically. Users can generate results using a specific taxonomy or run all three classifications simultaneously.

INTELLIGENCE Categorization Engine

With more than 800 entries, the Intelligence categorization engine provides wide coverage of diverse information domains. Categories include:
Campaign Finance
Fuel-Air Explosives
Censorship
Companies Chemicals / Petrochemicals
Oil & Gas Companies / Oil & Gas Drilling
Religious Conflict

A unique feature of this engine is its inclusion of a wide range of diverse topics and issues (chemical weapons, political crisis, wildlife, economic measures, cinema, terrorist groups, automotive, personal investments, etc.) from open domains, such as those found on open sources, social networks, reports and messaging systems (email, sms, chat, etc.).

Content Disclaimer: The information provided in this document is not to be reproduced and/or distributed without company consent.

CRIME & OFFENSE Categorization Engine

This taxonomy was developed to manage the information typical for the police and law enforcement domain. It includes a variety of categories such as:

Corruption
Murder, grievous bodily injury
Computer related crime

Computer Crime Categorization Engine

This taxonomy features a selection of specific categories to achieve accurate categorization of the Cyber Illegal domain. To master such a critical domain, we structured the taxonomy to include categories like DoS attack, Intrusion (computer or network), Identity theft, and others.

GEOTAXONOMY Categorization Engine

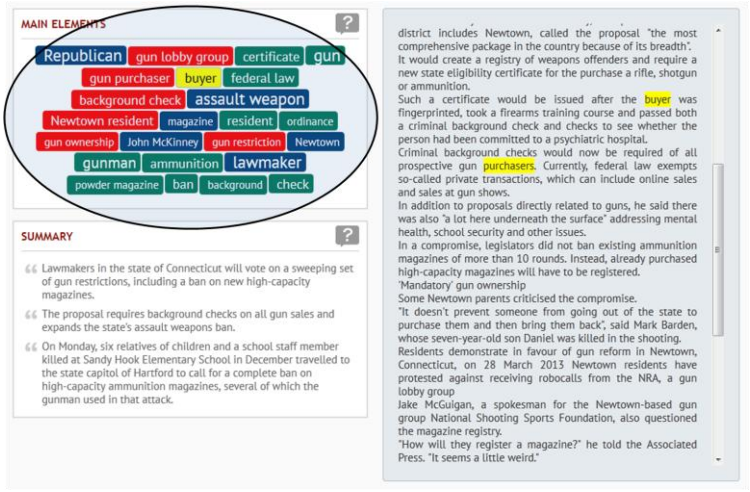
This classification engine produces output that highlights the countries and regions of the world in or associated with a text. Cogito uses a semantic approach to process geographic information present in text, deciphering ambiguities and correlating all existing concepts.

Tagging

Semantic processing of text produces three results that highlight:

- Document summary
- The most semantically relevant words and concepts
- Collocations that characterize the language text

Semantic analysis example:



“ Lawmakers in the state of Connecticut will vote on a sweeping set of gun restrictions, including a ban on new high-capacity magazines.

“ The proposal requires background checks on all gun sales and expands the state's assault weapons ban.

“ On Monday, six relatives of children and a school staff member killed at Sandy Hook Elementary School in December travelled to the state capitol of Hartford to call for a complete ban on high-capacity ammunition magazines, several of which the gunman used in that attack.

district includes Newtown, called the proposal "the most comprehensive package in the country because of its breadth".
It would create a registry of weapons offenders and require a new state eligibility certificate for the purchase a rifle, shotgun or ammunition.
Such a certificate would be issued after the buyer was fingerprinted, took a firearms training course and passed both a criminal background check and checks to see whether the person had been committed to a psychiatric hospital.
Criminal background checks would now be required of all prospective gun purchasers. Currently, federal law exempts so-called private transactions, which can include online sales and sales at gun shows.
In addition to proposals directly related to guns, he said there was also "a lot here underneath the surface" addressing mental health, school security and other issues.
In a compromise, legislators did not ban existing ammunition magazines of more than 10 rounds. Instead, already purchased high-capacity magazines will have to be registered.
'Mandatory' gun ownership
Some Newtown parents criticised the compromise.
"It doesn't prevent someone from going out of the state to purchase them and then bring them back", said Mark Barden, whose seven-year-old son Daniel was killed in the shooting.
Residents demonstrate in favour of gun reform in Newtown, Connecticut, on 28 March 2013 Newtown residents have protested against receiving robocalls from the NRA, a gun lobby group
Jake McGuigan, a spokesman for the Newtown-based gun group National Shooting Sports Foundation, also questioned the magazine registry.
"How will they register a magazine?" he told the Associated Press. "It seems a little weird."

Content Disclaimer: The information provided in this document is not to be reproduced and/or distributed without company consent.

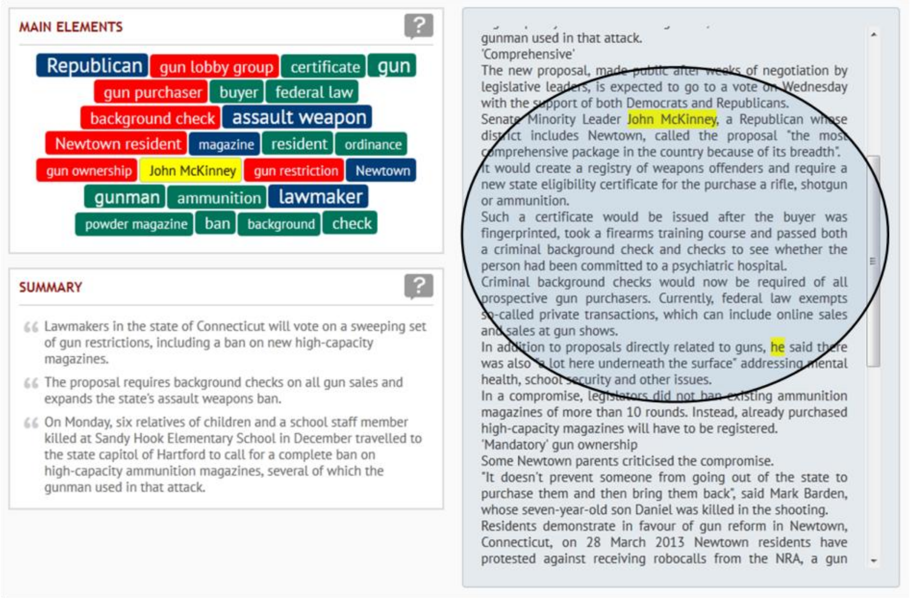
The results in the circled area above highlight:

- Lemmas or headwords (in *blue*)
- Concepts (in *green*)
- Semantically relevant collocations (in *red*)

Clicking on each element of the tag cloud highlights the element in the corresponding text. For example, clicking on the **concept** "OFFER" highlights the corresponding concept, in this case the synonym "PROPOSAL".

The same result is achieved by selecting any item from the tag cloud and/or phrases that are identified as the most significant (*see the box above the tag cloud*).

Selecting a person mentioned in the tag cloud will prompt the indicated name and personal pronoun, as shown in the following example where the proper name, "John McKinney" and "he" are highlighted.



MAIN ELEMENTS

SUMMARY

" Lawmakers in the state of Connecticut will vote on a sweeping set of gun restrictions, including a ban on new high-capacity magazines.

" The proposal requires background checks on all gun sales and expands the state's assault weapons ban.

" On Monday, six relatives of children and a school staff member killed at Sandy Hook Elementary School in December travelled to the state capitol of Hartford to call for a complete ban on high-capacity ammunition magazines, several of which the gunman used in that attack.

gunman used in that attack.
'Comprehensive'
The new proposal, made public after weeks of negotiation by legislative leaders, is expected to go to a vote on Wednesday with the support of both Democrats and Republicans.
Senate Minority Leader John McKinney, a Republican whose district includes Newtown, called the proposal "the most comprehensive package in the country because of its breadth".
It would create a registry of weapons offenders and require a new state eligibility certificate for the purchase a rifle, shotgun or ammunition.
Such a certificate would be issued after the buyer was fingerprinted, took a firearms training course and passed both a criminal background check and checks to see whether the person had been committed to a psychiatric hospital.
Criminal background checks would now be required of all prospective gun purchasers. Currently, federal law exempts so-called private transactions, which can include online sales and sales at gun shows.
In addition to proposals directly related to guns, he said there was also "a lot here underneath the surface" addressing mental health, school security and other issues.
In a compromise, legislators did not ban existing ammunition magazines of more than 10 rounds. Instead, already purchased high-capacity magazines will have to be registered.
'Mandatory' gun ownership
Some Newtown parents criticised the compromise.
"It doesn't prevent someone from going out of the state to purchase them and then bring them back", said Mark Barden, whose seven-year-old son Daniel was killed in the shooting.
Residents demonstrate in favour of gun reform in Newtown, Connecticut, on 28 March 2013 Newtown residents have protested against receiving robocalls from the NRA, a gun

These elements allow you to mark the context in the document with semantic information which can subsequently be used as follows:

- Highlights the most important sentences in text through a text preview or summary.
- Orders the search results giving a higher ranking to documents with the most semantically relevant keywords or concepts.
- Proposes correlations between different texts on the basis of language collocation.

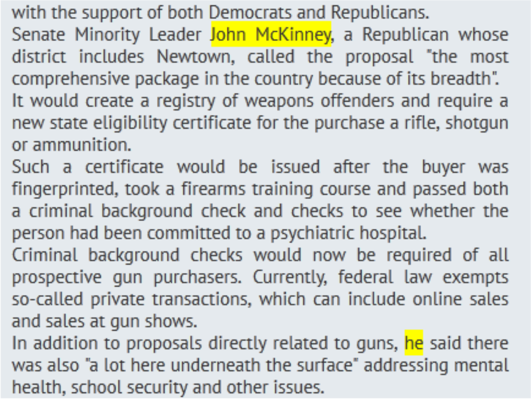
- Offers synthesized views in order to quickly and accurately grasp the weight and importance of the key information in the text.

Entity Extraction

The text mining entities are composed of **standard entities** and **domain entities**.

Cogito Intelligence API extracts a variety of standard entities including people, organization, places, dates, currencies, addresses, etc., even without the support of lists. The semantic engine is able to identify a proper name in text_and always correlate it back to its correct context. For example, "Arthur Andersen" may be categorized as “People” or “Organization” depending on the context.

In addition, to some extent, the anaphora that allows you to extract both the explicit and implicit references is also considered. The following example highlights the proper noun (People -> John McKinney), which relates to the personal pronoun "he" later in the text.



with the support of both Democrats and Republicans.
Senate Minority Leader John McKinney, a Republican whose district includes Newtown, called the proposal "the most comprehensive package in the country because of its breadth".
It would create a registry of weapons offenders and require a new state eligibility certificate for the purchase a rifle, shotgun or ammunition.
Such a certificate would be issued after the buyer was fingerprinted, took a firearms training course and passed both a criminal background check and checks to see whether the person had been committed to a psychiatric hospital.
Criminal background checks would now be required of all prospective gun purchasers. Currently, federal law exempts so-called private transactions, which can include online sales and sales at gun shows.
In addition to proposals directly related to guns, he said there was also "a lot here underneath the surface" addressing mental health, school security and other issues.

Domain entities may be defined as entities related to a specific context or realm of knowledge (Intelligence and LEA) which is regularly updated with current knowledge and events.

Cogito Intelligence API currently manages dozens of domain entities, including: Terrorist Organization, Biological Agents, World leaders, etc.

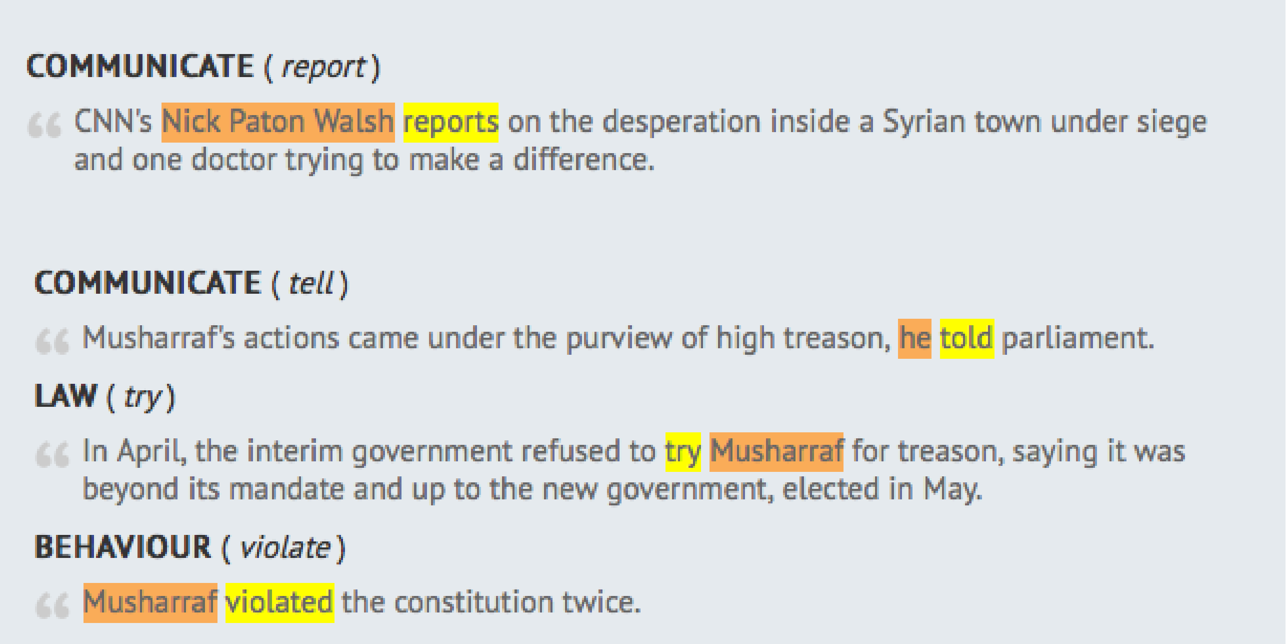
Thanks to the availability of customization tools (an option which can be acquired together with the training support Expert System provides to clients for special projects), clients can customize the extraction rules, enrich the semantic network or extract new entities based on their needs.

Relationships Extraction

In addition to the extraction of entities, **Cogito Intelligence API** can extract the *relationships between* semantic entities.

COGITO Intelligence API exceeds conventional entity extraction technology by offering coherent suggestions and contextualization of acquired text with reference to specific relations based on a set of over 20 different types of relationships.

Here are some examples:



COMMUNICATE (*report*)

“ CNN's Nick Paton Walsh reports on the desperation inside a Syrian town under siege and one doctor trying to make a difference.

COMMUNICATE (*tell*)

“ Musharraf's actions came under the purview of high treason, he told parliament.

LAW (*try*)

“ In April, the interim government refused to try Musharraf for treason, saying it was beyond its mandate and up to the new government, elected in May.

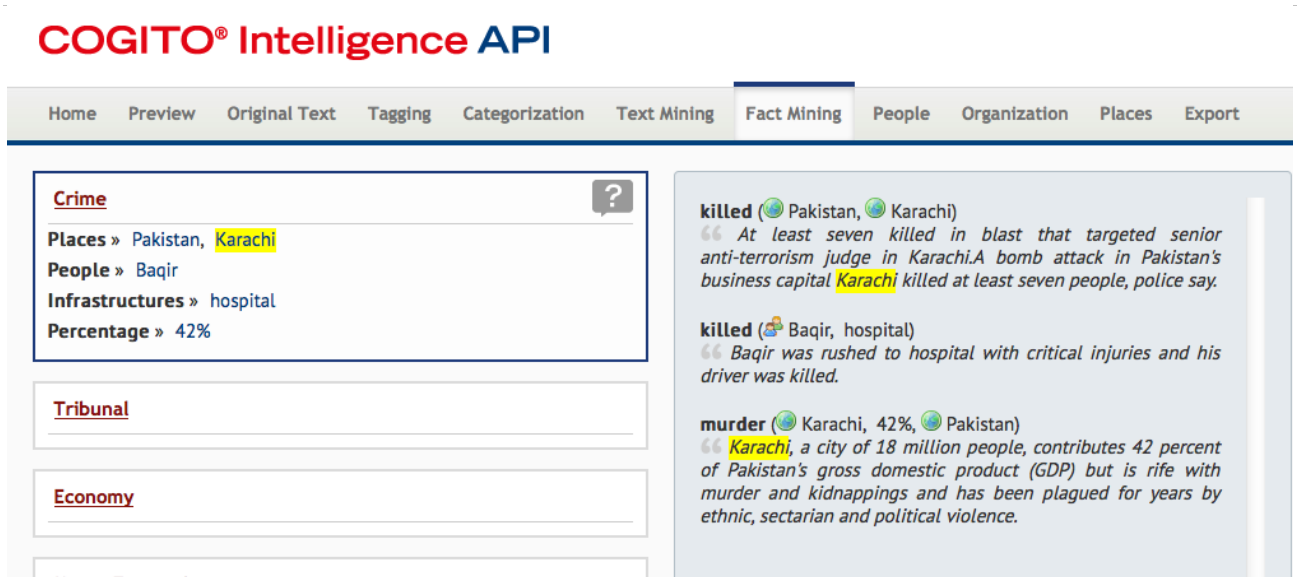
BEHAVIOUR (*violate*)

“ Musharraf violated the constitution twice.

Fact Mining

This innovative feature allows rapid identification of facts present in text, identifying both the fact, as well as the entities (people, organizations, places) and tags (URLs, phone numbers, emails, etc.) related to it. Both Intelligence and Crime taxonomies vehicle this information to trigger facts of interest and domain entities within.
In the example below, the words “killed” (and “murder”, “kidnapping”) identified in a fact (Crime) present in text, would generate a set of proposed entities (Bachir, Karachi, Pakistan, etc.) related to the fact.

Content Disclaimer: The information provided in this document is not to be reproduced and/or distributed without company consent.



Crime

Places » Pakistan, Karachi
People » Baqir
Infrastructures » hospital
Percentage » 42%

Tribunal

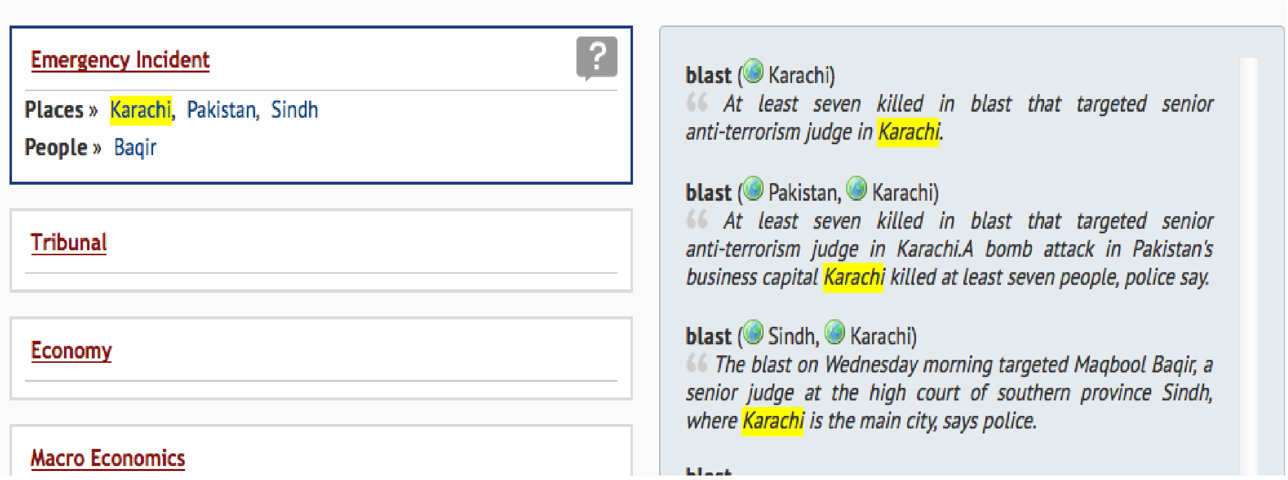
Economy

killed (Pakistan, Karachi)
“ At least seven killed in blast that targeted senior anti-terrorism judge in Karachi.A bomb attack in Pakistan's business capital Karachi killed at least seven people, police say.

killed (Baqir, hospital)
“ Baqir was rushed to hospital with critical injuries and his driver was killed.

murder (Karachi, 42%, Pakistan)
“ Karachi, a city of 18 million people, contributes 42 percent of Pakistan's gross domestic product (GDP) but is rife with murder and kidnappings and has been plagued for years by ethnic, sectarian and political violence.

In this second example you can see how the concept “blast” determines the fact “Emergency Incident” linked to Karachi, Pakistan, Sindh and Bakir.



Emergency Incident

Places » Karachi, Pakistan, Sindh
People » Baqir

Tribunal

Economy

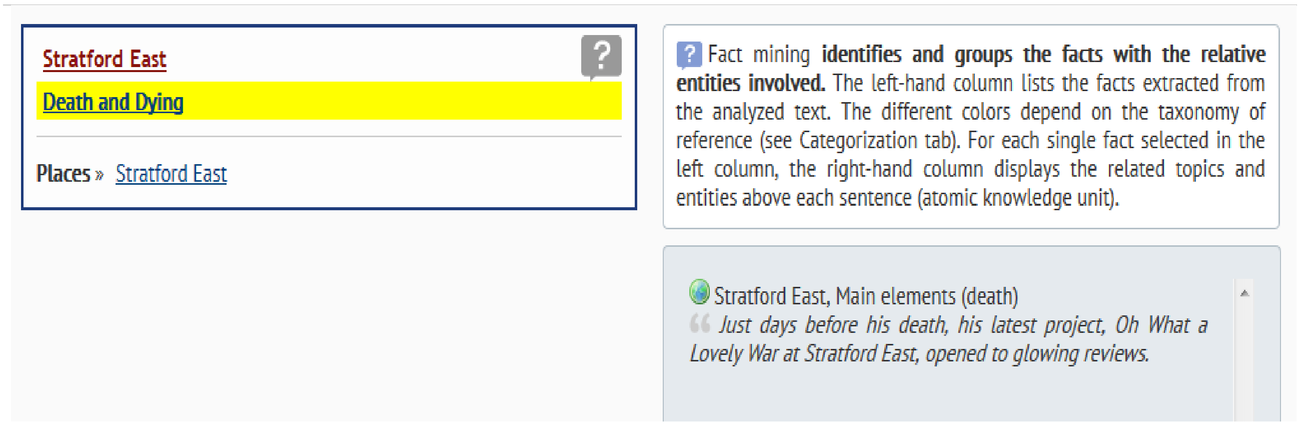
Macro Economics

blast (Karachi)
“ At least seven killed in blast that targeted senior anti-terrorism judge in Karachi.

blast (Pakistan, Karachi)
“ At least seven killed in blast that targeted senior anti-terrorism judge in Karachi.A bomb attack in Pakistan's business capital Karachi killed at least seven people, police say.

blast (Sindh, Karachi)
“ The blast on Wednesday morning targeted Maqbool Baqir, a senior judge at the high court of southern province Sindh, where Karachi is the main city, says police.

With a third “Geography” Fact Mining feature, it is possible to add a further perspective in gathering facts and contextualize them in the exact geographic location as detected in text.



Stratford East

Death and Dying

Places » Stratford East

Fact mining **identifies and groups the facts with the relative entities involved.** The left-hand column lists the facts extracted from the analyzed text. The different colors depend on the taxonomy of reference (see Categorization tab). For each single fact selected in the left column, the right-hand column displays the related topics and entities above each sentence (atomic knowledge unit).

Stratford East, Main elements (death)

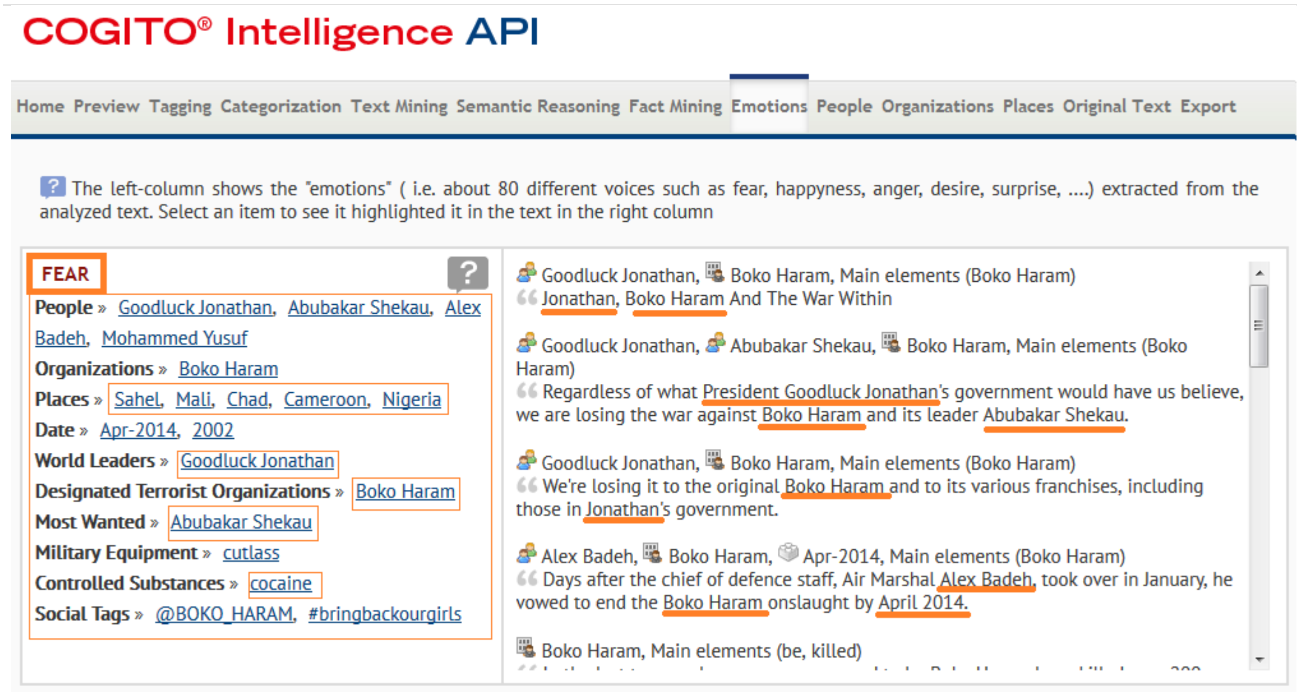
Just days before his death, his latest project, Oh What a Lovely War at Stratford East, opened to glowing reviews.

This change in facts' clusterization provides an innovative outlook from a geographic point of view.

Emotions

The Emotions feature masters about 80 categories making it possible to detect emotions within the text achieving a better and more focused emotions mining. This innovative feature breaks the boundaries of Sentiment Analysis transforming hidden emotions' content into accessible data and information which can be handled and associated to entities. The Emotions feature provides a new, innovative stage of text analysis.

The example below clearly shows how emotional contents of "fear" were found in the text pointing out all the sentences containing such information. The obtained data brings out a set of key entities (Abubakar Shekau, Boko Haram, Mali, cocaine, etc.) which can be associated to the Emotions' results providing a new analysis perspective.


COGITO® Intelligence API
Home Preview Tagging Categorization Text Mining Semantic Reasoning Fact Mining Emotions People Organizations Places Original Text Export

The left-column shows the "emotions" (i.e. about 80 different voices such as fear, happyness, anger, desire, surprise, ...) extracted from the analyzed text. Select an item to see it highlighted it in the text in the right column

FEAR

People » Goodluck Jonathan, Abubakar Shekau, Alex Badeh, Mohammed Yusuf
Organizations » Boko Haram
Places » Sahel, Mali, Chad, Cameroon, Nigeria
Date » Apr-2014, 2002
World Leaders » Goodluck Jonathan
Designated Terrorist Organizations » Boko Haram
Most Wanted » Abubakar Shekau
Military Equipment » cutlass
Controlled Substances » cocaine
Social Tags » @BOKO_HARAM, #bringbackourgirls

Goodluck Jonathan, Boko Haram, Main elements (Boko Haram)
“ Jonathan, Boko Haram And The War Within

Goodluck Jonathan, Abubakar Shekau, Boko Haram, Main elements (Boko Haram)
“ Regardless of what President Goodluck Jonathan's government would have us believe, we are losing the war against Boko Haram and its leader Abubakar Shekau.

Goodluck Jonathan, Boko Haram, Main elements (Boko Haram)
“ We're losing it to the original Boko Haram and to its various franchises, including those in Jonathan's government.

Alex Badeh, Boko Haram, Apr-2014, Main elements (Boko Haram)
“ Days after the chief of defence staff, Air Marshal Alex Badeh, took over in January, he vowed to end the Boko Haram onslaught by April 2014.

Boko Haram, Main elements (be, killed)

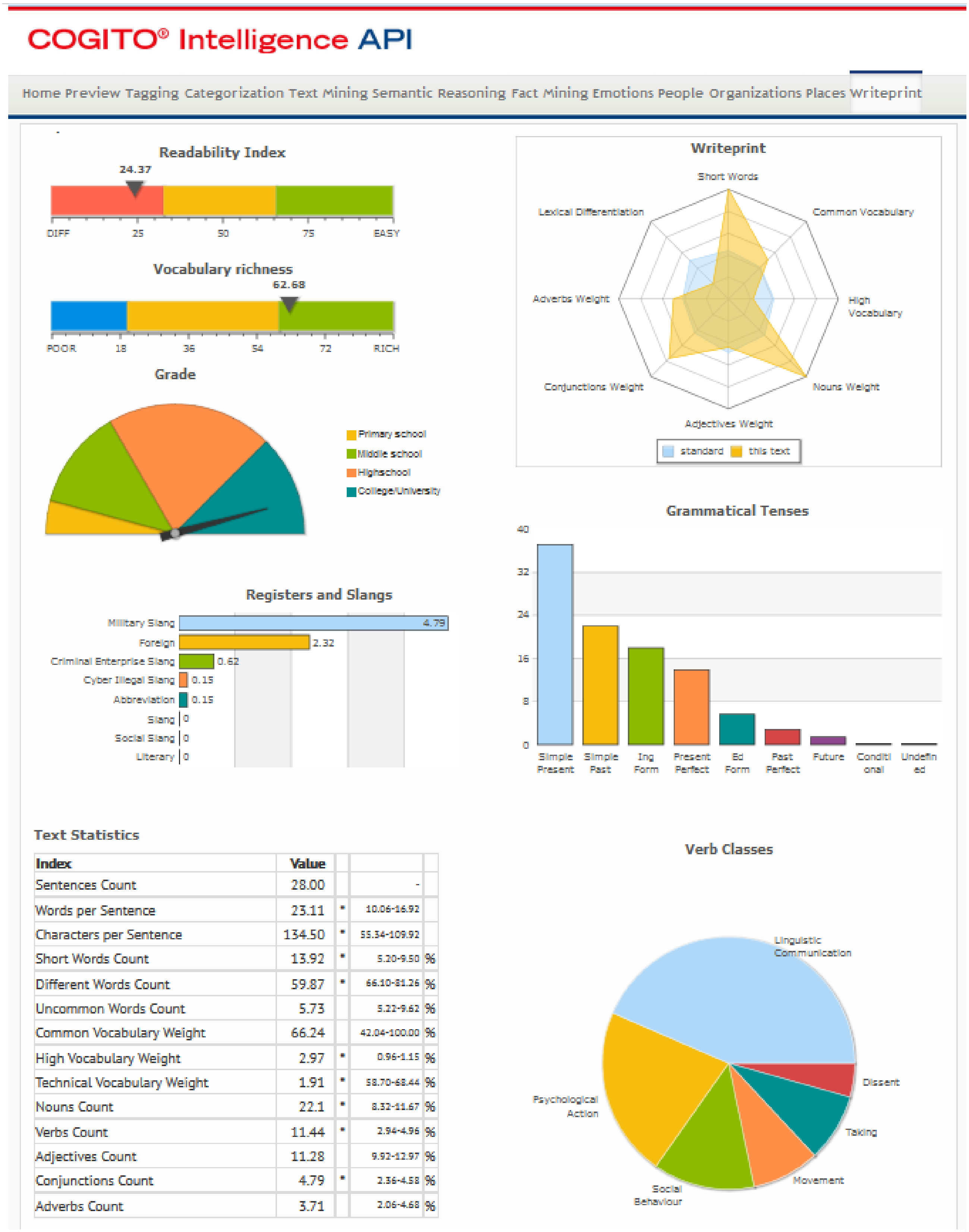
Writeprint

The Writeprint feature introduces a whole new level of language analysis providing powerful statistical and semantic text readability indexes to target Biometry and authorship assessment.
With more than 65 core indexes, Writeprint is capable of outlining a document’s readability level and the education grade necessary to understand it. The new feature also provides a full set of grammatical and structural analysis items.

Cogito Intelligence API’s readability index is based on standard algorithms for which there’s a wide range of scientific literature. We studied and reexamined the Coleman Liau index adding more of our semantic factor greatly increasing our indexes’ precision and reliability.
The feature is powered with two sets of statistical and semantic indexes. With more than 5 domain specific slangs and registers, it is possible to achieve better slangs disambiguation, writing styles and topics.

The example below clearly shows how Writeprint is capable of analyzing documents deeply into their textual and grammatical structures working the readability values out. On the right-hand column, the Writeprint graph compares the different scores obtained in the most crucial indexes leaving a “(write-)print” of the author’s style and peculiarities.

Content Disclaimer: The information provided in this document is not to be reproduced and/or distributed without company consent.

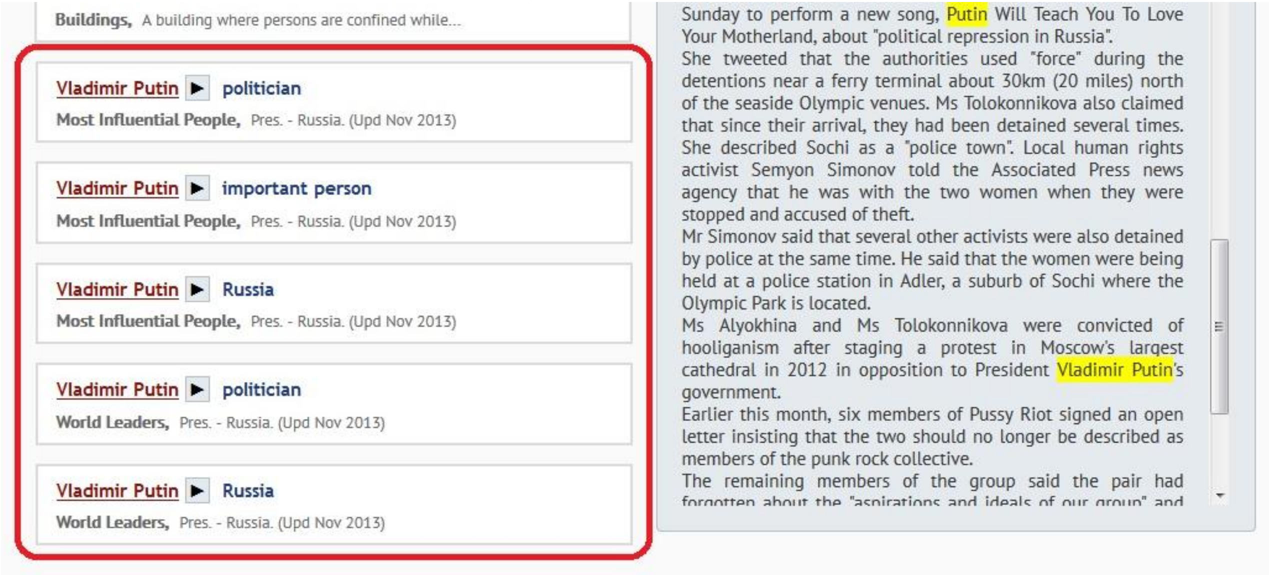


Text Statistics

Index	Value			
Sentences Count	28.00		-	
Words per Sentence	23.11	*	10.06-16.92	
Characters per Sentence	134.50	*	55.34-109.92	
Short Words Count	13.92	*	5.20-9.50	%
Different Words Count	59.87	*	66.10-81.26	%
Uncommon Words Count	5.73		5.22-9.62	%
Common Vocabulary Weight	66.24		42.04-100.00	%
High Vocabulary Weight	2.97	*	0.96-1.15	%
Technical Vocabulary Weight	1.91	*	58.70-68.44	%
Nouns Count	22.1	*	8.32-11.67	%
Verbs Count	11.44	*	2.94-4.96	%
Adjectives Count	11.28		9.92-12.97	%
Conjunctions Count	4.79	*	2.36-4.58	%
Adverbs Count	3.71		2.06-4.68	%

Semantic Reasoning

Semantic reasoning brings a new, innovative function which extends the Text mining feature. In fact, for entities related to the Intelligence and Security domain, Semantic Reasoning is able to automatically infer information NOT present within the text, thus providing consistent information about domain entities. As shown in the example below, the Reasoning triggers a process of consistent information enrichment which supports inferences formulation and data analysis.



Buildings, A building where persons are confined while...

Vladimir Putin ▶ **politician**
Most Influential People, Pres. - Russia. (Upd Nov 2013)

Vladimir Putin ▶ **important person**
Most Influential People, Pres. - Russia. (Upd Nov 2013)

Vladimir Putin ▶ **Russia**
Most Influential People, Pres. - Russia. (Upd Nov 2013)

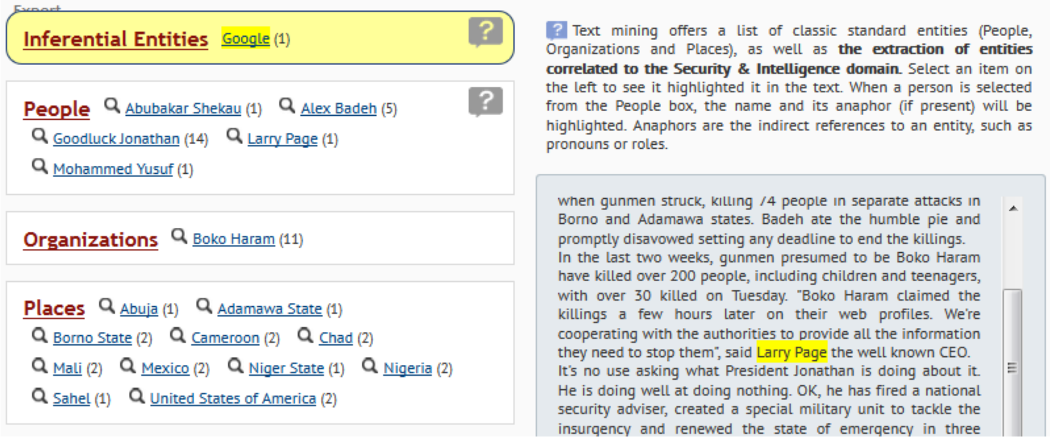
Vladimir Putin ▶ **politician**
World Leaders, Pres. - Russia. (Upd Nov 2013)

Vladimir Putin ▶ **Russia**
World Leaders, Pres. - Russia. (Upd Nov 2013)

Sunday to perform a new song, Putin Will Teach You To Love Your Motherland, about "political repression in Russia".
She tweeted that the authorities used "force" during the detentions near a ferry terminal about 30km (20 miles) north of the seaside Olympic venues. Ms Tolokonnikova also claimed that since their arrival, they had been detained several times. She described Sochi as a "police town". Local human rights activist Semyon Simonov told the Associated Press news agency that he was with the two women when they were stopped and accused of theft.
Mr Simonov said that several other activists were also detained by police at the same time. He said that the women were being held at a police station in Adler, a suburb of Sochi where the Olympic Park is located.
Ms Alyokhina and Ms Tolokonnikova were convicted of hooliganism after staging a protest in Moscow's largest cathedral in 2012 in opposition to President Vladimir Putin's government.
Earlier this month, six members of Pussy Riot signed an open letter insisting that the two should no longer be described as members of the punk rock collective.
The remaining members of the group said the pair had forgotten about the "aspirations and ideals of our group" and

Inferential Entities

Inferential Entities are unveiled by assumption of their strong connection with the entities in the text, as highlighted by Semantic Reasoning. So the inferred information is not taken from text but is strongly consistent and related to the detected entities.



Export

Inferential Entities Google (1)

People Abubakar Shekau (1) Alex Badeh (5) Goodluck Jonathan (14) Larry Page (1) Mohammed Yusuf (1)

Organizations Boko Haram (11)

Places Abuja (1) Adamawa State (1) Borno State (2) Cameroon (2) Chad (2) Mali (2) Mexico (2) Niger State (1) Nigeria (2) Sahel (1) United States of America (2)

Text mining offers a list of classic standard entities (People, Organizations and Places), as well as **the extraction of entities correlated to the Security & Intelligence domain**. Select an item on the left to see it highlighted it in the text. When a person is selected from the People box, the name and its anaphor (if present) will be highlighted. Anaphors are the indirect references to an entity, such as pronouns or roles.

when gunmen struck, killing 74 people in separate attacks in Borno and Adamawa states. Badeh ate the humble pie and promptly disavowed setting any deadline to end the killings.
In the last two weeks, gunmen presumed to be Boko Haram have killed over 200 people, including children and teenagers, with over 30 killed on Tuesday. "Boko Haram claimed the killings a few hours later on their web profiles. We're cooperating with the authorities to provide all the information they need to stop them", said Larry Page the well known CEO.
It's no use asking what President Jonathan is doing about it. He is doing well at doing nothing. OK, he has fired a national security adviser, created a special military unit to tackle the insurgency and renewed the state of emergency in three

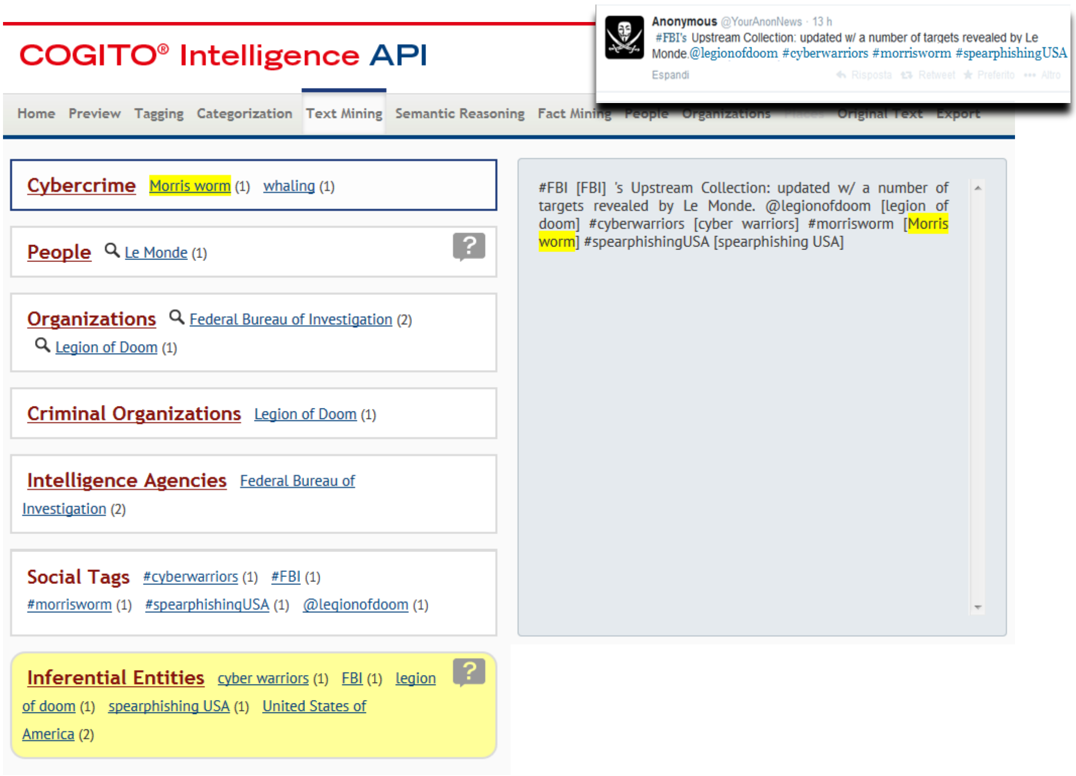
Inferential Entities apply to all People, Organizations and Places entities providing a higher abstraction level to extract further information from inferred metadata and picture new semantic overviews.

Social Tags Normalization

This innovative feature allows immediate detection of social and web related data, and processes the information to retrieve its meaning. Our Social Tags Normalization feature is the first tool to actually gather genuine information from social networks by detecting crucial entities in hashtags, profile names and URLs.

@syrianelectronicarmy → Cogito turns it into → Syrian electronic army
@_cypherpunks_ → Cogito turns it into → cypher punks
@alqaeda → Cogito turns it into → al qaeda
#bringbackourgirls → Cogito turns it into → bring back our girls
#narcoterrorism → Cogito turns it into → narco terrorism
#elections2014 → Cogito turns it into → elections 2014

All of the highlighted texts are core entities which are recognized and extracted from social data. That would not be possible without our normalization feature.



Anonymous @YourAnonNews · 13 h
#FBI's Upstream Collection: updated w/ a number of targets revealed by Le Monde.@legionofdoom #cyberwarriors #morrisworm #spearphishingUSA
Espandi

Cybercrime Morris worm (1) whaling (1)

People Le Monde (1)

Organizations Federal Bureau of Investigation (2) Legion of Doom (1)

Criminal Organizations Legion of Doom (1)

Intelligence Agencies Federal Bureau of Investigation (2)

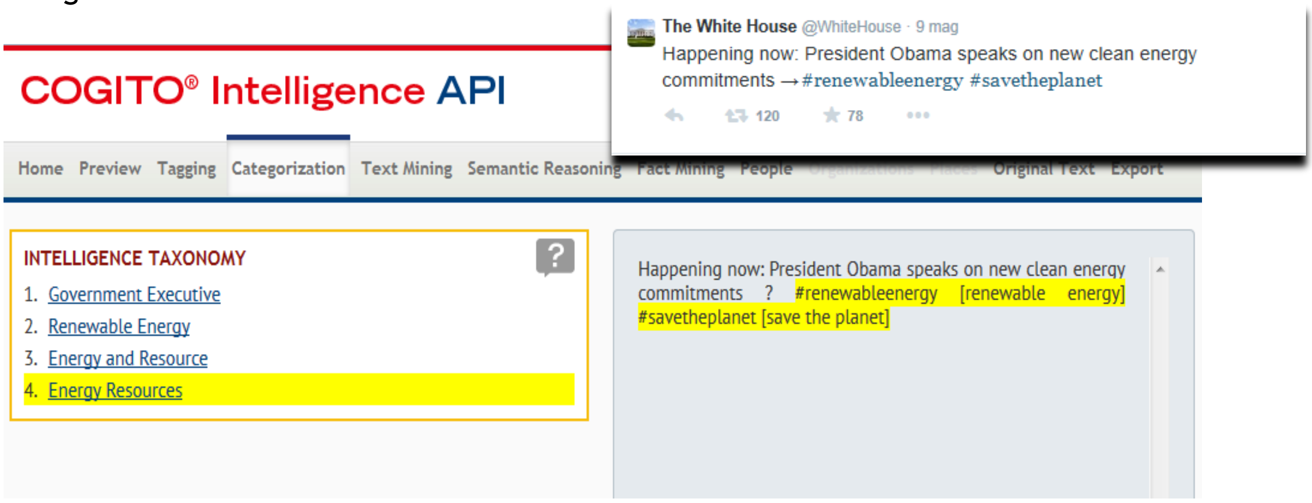
Social Tags #cyberwarriors (1) #FBI (1) #morrisworm (1) #spearphishingUSA (1) @legionofdoom (1)

Inferential Entities cyber warriors (1) FBI (1) legion of doom (1) spearphishing USA (1) United States of America (2)

#FBI [FBI] 's Upstream Collection: updated w/ a number of targets revealed by Le Monde. @legionofdoom [legion of doom] #cyberwarriors [cyber warriors] #morrisworm [Morris worm] #spearphishingUSA [spearphishing USA]

Content Disclaimer: The information provided in this document is not to be reproduced and/or distributed without company consent.

Normalization also provides crucial information to achieve precise semantic categorization.



The White House @WhiteHouse · 9 mag
Happening now: President Obama speaks on new clean energy commitments → #renewableenergy #savetheplanet

INTELLIGENCE TAXONOMY

1. Government Executive
2. Renewable Energy
3. Energy and Resource
4. Energy Resources

Happening now: President Obama speaks on new clean energy commitments ? #renewableenergy [renewable energy] #savetheplanet [save the planet]

Content Disclaimer: The information provided in this document is not to be reproduced and/or distributed without company consent.