THOMSON REUTERS OPEN CALAIS

API USER GUIDE

THOMSON REUTERS

Contents

Chapter 1   Open Calais

1.1   What is Open Calais?

Open Calais is a sophisticated Thomson Reuters web service that attaches intelligent metadata-tags to your unstructured content, enabling powerful text analytics. The Open Calais natural language processing engine automatically analyzes and tags your input files in such a way that your consuming application can both easily pinpoint relevant data, and effectively leverage the invaluable intelligence and insights contained within the text.

Open Calais analyzes the semantic content of your input files using a combination of statistical, machine-learning, and custom pattern-based methods. Developed by the Text Metadata Services (TMS) group at Thomson Reuters, Open Calais outputs highly accurate and detailed metadata.

Open Calais also maps your metadata-tags to Thomson Reuters unique IDs. This supports disambiguation (and linking) of data across all documents processed by Open Calais, and also offers you the opportunity to further enrich your data with related information from the Thomson Reuters datasets.

1.2   How Does Open Calais Work?

Open Calais automatically analyzes your input text and performs the following processes:

- Named Entity and Relationship Recognition – Open Calais identifies and tags mentions (text strings) of things like companies, people, deals, geographical locations, industries, physical assets, organizations, products, events, etc., based on a list of predefined metadata types.

- Aboutness Tagging – Open Calais assigns social, topic, and industry tags that describe what the input document is about as a whole.

1.2.1   Named Entity and Relationship Recognition

During processing, Open Calais automatically scans and analyzes the input text, searching for mentions of things like companies, people, cities, industries, products, deals, alliances, company earnings announcements, company layoffs, IPOs, stock splits, business relationships, etc.

Open Calais classifies mentions of straightforward things like companies, people, cities, telephone numbers, etc. as **Entities;** more complex mentions that indicate relationships between things are classified as **Relations**. Some examples of relations are: deals, IPOs, analyst recommendations, company reorganizations, product recalls.

For the complete list of Open Calais Entity and Relation types, see Entity Markup Tags.

Open Calais outputs the following named entity and relationship tags:

- Instance Tags– Each mention found by Open Calais is expressed as an Instance tag.

- Entity Markup Tag– Each group of one or more instances deemed to refer to a unique thing is expressed as an Entity Markup tag. (For example, multiple mentions of the same person will generate a single Entity Markup tag of the type Person; multiple mentions of the same company will generate a single Entity Markup tag of the type Company; multiple mentions of the same deal will generate a single Entity Markup tag of the type Deal; etc.) This is what we call the "**extracted entity**" or the "**extracted relation**."

- Relevance Tag– A tag that indicates how centric the extracted entity or relation is to the containing document.

- Confidence Tag– A tag that indicates how confident we are that the extracted entity of the type company, for example, is indeed a company, or how confident we are that the extracted entity of the type person, for example, is indeed a person.

- Disambiguation Tag– Open Calais attempts to map an extracted entity or relation to the corresponding entity and unique ID in the relevant Thomson Reuters dataset. If the mapping is successful, a Disambiguation tag is generated. The mapping is what enables all the instances, extracted entities, and extracted relations that refer to the same thing to be unambiguously identified (and thus linked) across all documents processed by Open Calais.

1.2.1.1   Instance Tags

Each mention of a predefined entity or relation type found by Open Calais is expressed as an Instance tag in the output file. The Instance tag describes the mention. It includes the "found" text string itself, the surrounding text, the location and offset of the text string. Each instance is assigned a unique ID.

For example, Open Calais found the following mentions of Tim Cook, the CEO of Apple, Inc., in an article about the anticipated launch of the Apple Watch:

"All Eyes on Apple's **Cook** as Watch Launch Expected"

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e/Instance
/11">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/InstanceInfo"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e"/>
        <c:subject rdf:resource="http://d.opencalais.com/pershash-1/e4808181-2cd0-3670-b992-7467229ba691"/>
        <!--Person: Tim Cook; -->
        <c:detection>[&lt;Title&gt;All Eyes on Apple's ]Cook[ as Watch Launch
        Expected&lt;/Title&gt;]</c:detection>
        <c:prefix>&lt;Title&gt;All Eyes on Apple's </c:prefix>
        <c:exact>Cook</c:exact>
        <c:suffix> as Watch Launch Expected&lt;/Title&gt;</c:suffix>
        <c:offset>40</c:offset>
        <c:length>4</c:length>
    </rdf:Description>
```

InstanceInfo Tag Attributes

"Apple Inc Chief Executive Officer **Tim Cook** on Monday is expected to announce details of the first product developed under his leadership, a watch that Apple hopes will transform the market of wearable technology."

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e/Instance/12">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/InstanceInfo"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e"/>
        <c:subject rdf:resource="http://d.opencalais.com/pershash-1/e4808181-2cd0-3670-b992-7467229ba691"/>
        <!--Person: Tim Cook; -->
        <c:detection>[ (Reuters) - Apple Inc Chief Executive Officer ]Tim Cook[ on Monday is expected to
        announce details of the]</c:detection>
        <c:prefix> (Reuters) - Apple Inc Chief Executive Officer </c:prefix>
        <c:exact>Tim Cook</c:exact>
        <c:suffix> on Monday is expected to announce details of the</c:suffix>
        <c:offset>162</c:offset>
        <c:length>8</c:length>
    </rdf:Description>
```

InstanceInfo Tag Attributes

"Apple Inc Chief Executive Officer Tim Cook on Monday is expected to announce details of the first product developed under **his** leadership, a watch that Apple hopes will transform the market of wearable technology."

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e/Instance/13">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/InstanceInfo"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e"/>
        <c:subject rdf:resource="http://d.opencalais.com/pershash-1/e4808181-2cd0-3670-b992-7467229ba691"/>
        <!--Person: Tim Cook; -->
        <c:detection>[details of the first product developed under ]his[ leadership, a watch that Apple hopes
will]</c:detection>
        <c:prefix>details of the first product developed under </c:prefix>
        <c:exact>his</c:exact>
        <c:suffix> leadership, a watch that Apple hopes will</c:suffix>
        <c:offset>250</c:offset>
        <c:length>3</c:length>
    </rdf:Description>
```

InstanceInfo Tag Attributes

"Apple will have to 'tweak' its stores to handle the watch, **Cook** told the Telegraph newspaper recently."

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e/Instance/14">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/InstanceInfo"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e"/>
        <c:subject rdf:resource="http://d.opencalais.com/pershash-1/e4808181-2cd0-3670-b992-7467229ba691"/>
        <!--Person: Tim Cook; -->
        <c:detection>[have to "tweak" its stores to handle the watch, ]Cook[ told the Telegraph newspaper
recently. Cook]</c:detection>
        <c:prefix>have to "tweak" its stores to handle the watch, </c:prefix>
        <c:exact>Cook</c:exact>
        <c:suffix> told the Telegraph newspaper recently. Cook</c:suffix>
        <c:offset>2038</c:offset>
        <c:length>4</c:length>
    </rdf:Description>
```

InstanceInfo Tag Attributes

1.2.1.2   Entity Markup Tags

Open Calais identifies the instances that refer to the same thing and links them to each other. Each group of instances deemed to refer to a unique thing (e.g. one or more mentions of the same company, or one or more mentions of the same person or the same deal) results in a single **Entity Markup tag** in the output file. This is what we call the "extracted entity" or the "extracted relation."

Note that the original mentions do not have to be identical text strings in order to be recognized as referring to the same thing. For example, Tim Cook, may be referred to as "Tim Cook", "Timothy Cook," "Timothy Donald Cook," "Mr. Cook," "Cook," and even "he," or "his." Each mention that refers to Tim Cook is expressed as an Instance in the output file; Open Calais identifies the instances that refer to the same thing (in this case, Tim Cook), and outputs a single extracted entity or relation, expressed in the output file as an entity markup tag.

The em/e/person tag extracted for Tim Cook:

```
<rdf:Description rdf:about="http://d.opencalais.com/pershash-1/e4808181-2cd0-3670-b992-7467229ba691">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/em/e/Person"/>
        <c:permid rdf:resource="https://permid.org/1-404011"/>
        <c:forenduserdisplay>true</c:forenduserdisplay>
        <c:name>Tim Cook</c:name>
        <c:persontype>economic</c:persontype>
        <c:nationality>N/A</c:nationality>
        <c:confidencelevel>0.999</c:confidencelevel>
        <c:commonname>Tim Cook</c:commonname>
    </rdf:Description>
```

[em/e/Person Tag Attributes](#)

Every entity markup tag has one or more related instances.

Open Calais assigns a unique ID (a hash tag) to the extracted entity. This ID also appears in the related instances, linking them to the extracted entity and to each other. In this example, the hash tag for the extracted entity, Tim Cook, is *pershash-1/e4808181-2cd0-3670-b992-7467229ba691*. Note that the same hash tag appears in the instance tags generated for mentions of Tim Cook.

For the complete list of Open Calais Entity and Relation types, and a description of each see [Entity Markup Tags](#).

1.2.1.3   Relevance Tags

All extracted entities have an associated Relevance tag that indicates how centric the entity is to the containing document. Relevance scores range from 0 to 1. The higher the score, the more relevant the entity is to the containing document.

The subject attribute of the Relevance Tag indicates the entity that is the subject of the relevance score. In the following example, the subject attribute value is the hash tag that points to the extracted person entity, Tim Cook. The high relevance score indicates that the person Tim Cook is indeed centric to this story.

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e/Relevance/9">
      <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/RelevanceInfo"/>
      <c:docId rdf:resource="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e"/>
      <c:subject rdf:resource="http://d.opencalais.com/pershash-1/e4808181-2cd0-3670-b992-7467229ba691"/>
      <c:relevance>0.8</c:relevance>
   </rdf:Description>
```

RelevanceInfo Tag Attributes

1.2.1.4   Confidence Tags

Intensive efforts are devoted to making tagging as accurate as possible; however, automated tagging will never be 100% accurate. Therefore Open Calais implements confidence scoring. Confidence scoring indicates the likelihood that the extracted e.g. person or company is indeed a person or company.

The extracted entities and relations that implement confidence scores display the **confidencelevel** attribute within the Entity Markup tag, and may also generate a related Confidence tag. The higher the confidence score, the more confident we are that the e.g. extracted person or company is indeed a person or company.

The consuming application can use the confidence score to achieve higher accuracy results by ignoring entities and relations and their related tags with confidence scores below a specified level. Note that when you raise the specified level, you are boosting precision at the expense of recall, increasing the risk of ignoring tags that are correct. If you choose to filter data based on this feature, you should adjust the confidence threshold according to the specific use case.

In the following example, the high confidence score, 0.999 indicates a high likelihood that the extracted entity of the type person is indeed a person. Once again, note the hash tag (pershash) linking this Confidence tag to the em/e/person tag for Tim Cook.

```
<rdf:Description rdf:about="http://d.opencalais.com/conf/pershash-1/e4808181-2cd0-3670-b992-7467229ba691">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/Confidence"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e"/>
        <!--Tim Cook-->
        <c:subject rdf:resource="http://d.opencalais.com/pershash-1/e4808181-2cd0-3670-b992-7467229ba691"/>
        <c:statisticalfeature>0.999</c:statisticalfeature>
        <c:dblookup>0.95</c:dblookup>
        <c:resolution>0.0</c:resolution>
        <c:aggregate>0.999</c:aggregate>
    </rdf:Description>
```

Confidence Tag Attributes

1.2.1.5   Disambiguation Tags

Open Calais attempts to map extracted entities and relations to the corresponding entities and unique IDs in the relevant Thomson Reuters dataset. If the mapping is successful, a **Disambiguation tag** is created in the output file. The linking to a Thomson Reuters unique ID is an exact and specific identity recognition. The mapping is what enables all the instances, extracted entities, and extracted relations that refer to the same thing to be unambiguously identified (and thus linked) across all documents processed by Open Calais.

Additionally, the mapping offers you the opportunity to further enrich your data with information from the Thomson Reuters datasets. For further information about how to leverage the Thomson Reuters IDs, browse to https://permid.org.

The Disambiguation tag (er/person) extracted for Tim Cook:

```
<rdf:Description rdf:about="http://d.opencalais.com/er/person/ralg-pa1/34413199178">
    <rdf:type rdf:resource="http://s.opencalais.com/1/type/er/Person"/>
    <c:docId rdf:resource="http://d.opencalais.com/ dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e"/>
    <c:name>Timothy D. Cook</c:name>
    <c:personid>88090</c:personid>
    <c:paid>34413199178</c:paid>
    <c:officerid>88090</c:officerid>
    <c:commonname>Tim Cook</c:commonname>
    <c:score>0.9358713</c:score>
    <!--Tim Cook-->
    <c:subject rdf:resource="http://d.opencalais.com/pershash-1/e4808181-2cd0-3670-b992-7467229ba691"/>
    <c:openpermid rdf:resource="https://permid.org/1-4295905573"/>
</rdf:Description>
```

er/Person Tag Attributes

The hash tag generated by Open Calais is a local ID that links extracted entities or relations with their related tags within the containing document (local disambiguation), while the Thomson Reuters unique ID (the <c:paid> attribute value in this example) displayed in the ER tag will be consistent across all documents processed by Open Calais. (Note that the *paid* is equivalent to a *permid*.)

For a description of each of the Open Calais Disambiguation types see Disambiguation Tags.

Note: The **openpermid** attribute of the Person, Company, and TopmostPublicParentCompany disambiguation types gives you direct access to high quality, curated Thomson Reuters company data. The attribute value is a direct link to the relevant company page on the Open PermID website (https://permid.org).

1.2.2   Aboutness Tagging

In addition to identifying and tagging individual text strings, Open Calais further enriches your data with metadata tags designed to describe the piece of content as a whole:

- Social Tagging – Classifies the document based on Wikipedia folksonomy.

- Topic Tagging – Identifies the topics discussed in the document. The list of possible topics is defined by the Thomson Reuters Coding Schema (TRCS) and International Press Telecommunications Council (IPTC) taxonomies.

- Industry Tagging –Identifies the industries related to the text. The list of industries that can be identified is defined by the Thomson Reuters Business Classification (TRBC) taxonomy.

1.2.2.1   Social Tags

A Social Tag is an association of the submitted text to related Wikipedia categories, or articles. Social tags attempt to emulate how a person would tag a specific piece of content. For example, if you submit a story about Barak Obama and a piece of legislation, at least one reasonable tag would be "U.S. Legislation." A story about the relative merits of BMWs, Ferraris, and Porsches would probably be tagged with "sports cars," "luxury makes," "auto racing," and "motorsport."

The story about the Apple Watch Launch generated the following social tags: IOS, Smartwatches, Wearable Computers, Human-computer interaction, Ubiquitous computing, Consumer electronics, Apple Inc., Wearable Technology, and Apple system on a chip.

The SocialTag function does not identify individual items within the text, but rather attempts to provide common sense tags for the piece of content as a whole.

Social tags are derived from the Wikipedia folksonomy. They are periodically updated to keep them current.

Examples of social tags:

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e/SocialTag/1">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/SocialTag"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e"/>
        <c:socialtag rdf:resource="http://d.opencalais.com/genericHasher-1/93511881-ab5b-3bd5-8f7b-
        c12bb17d2f38"/>
        <c:forenduserdisplay>true</c:forenduserdisplay>
        <c:name>IOS</c:name>
        <c:importance>1</c:importance>
        <c:originalValue>IOS (Apple)</c:originalValue>
</rdf:Description>
```

SocialTag Attributes


```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e/SocialTag/2">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/SocialTag"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e"/>
        <c:socialtag rdf:resource="http://d.opencalais.com/genericHasher-1/bcb1987a-9543-3dd8-a970-
        2a0229480c58"/>
        <c:forenduserdisplay>true</c:forenduserdisplay>
        <c:name>Smartwatches</c:name>
        <c:importance>1</c:importance>
        <c:originalValue>Smartwatches</c:originalValue>
</rdf:Description>
```

SocialTag Attributes


```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e/SocialTag/3">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/SocialTag"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e"/>
        <c:socialtag rdf:resource="http://d.opencalais.com/genericHasher-1/68a40065-c128-3179-89ad-
        5f102c21f926"/>
        <c:forenduserdisplay>true</c:forenduserdisplay>
        <c:name>Wearable computers</c:name>
        <c:importance>1</c:importance>
        <c:originalValue>Wearable computers</c:originalValue>
    </rdf:Description>
```

SocialTag Attributes

1.2.2.2   Topic Tags (DocCat)

Open Calais identifies the topic or topics that are being discussed in the document. For example, "Macroeconomics," "Equities," "Sports," "Entertainment," "Politics," "Oil & Gas Products," "Mergers/Acquisitions/Takeovers," "Computer Hardware," "Consumer Financial Services," "Software and IT Services," etc. A DocCat (topic) tag is designed to give a general notion of what an input document is about. There is no specific entity recognition in the text, but rather deduction about what the text is about.

The reference list of topics is defined by the Thomson Reuters Coding Schema (TRCS) and/or by the International Press Telecommunications Council (IPTC) news taxonomy and/or by the Self Service Classification project taxonomy.

Each identified topic results in a Topic (DocCat) tag. It is possible that multiple topics will be identified, or that no topic will be identified if the content does not discuss anything currently defined by the relevant taxonomies.

Following are some of the Topic tags that were extracted by Open Calais from the story about the Apple Watch Launch.

An IPTC taxonomy topic:

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e/cat/1">
       <rdf:type rdf:resource="http://s.opencalais.com/1/type/cat/DocCat"/>
       <c:docId rdf:resource="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e"/>
       <c:forenduserdisplay>false</c:forenduserdisplay>
       <c:score>0.982</c:score>
       <c:name>Technology_Internet</c:name>
    </rdf:Description>
```

[Topic Tag (DocCat tag) attributes](#)

A TRCS taxonomy topic:

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e/cat/2">
       <rdf:type rdf:resource="http://s.opencalais.com/1/type/cat/DocCat"/>
       <c:docId rdf:resource="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e"/>
       <c:forenduserdisplay>false</c:forenduserdisplay>
       <c:rcscode>B:162</c:rcscode>
       <c:name>Technology Equipment</c:name>
       <c:shortName>Technology Equipment</c:shortName>
       <c:score>0.273</c:score>
    </rdf:Description>
```

[Topic Tag (DocCat tag) attributes](#)

1.2.2.3   Industry Tags

During processing, Open Calais identifies the industries that are related to the companies mentioned in the text. For example, "Management Consultant Services," "Information Services," "Biotechnology & Medical Services," "Integrated Telecommunications Services – NEC," "Handbags and Luggage Retailers," "Petroleum Refining," etc.

The list of industries that can be identified is defined by the Thomson Reuters Business Classification (TRBC) taxonomy.

Industry Tags include a unique Thomson Reuters ID. This ID enables extracting information about the industry from the Thomson Reuters dataset, and also supports linkage across documents processed by Open Calais.

The following Industry tags were extracted by Open Calais from the story about the Apple Watch Launch:

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e/Industry/2">
      <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/Industry"/>
      <c:docId rdf:resource="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e"/>
      <c:forenduserdisplay>false</c:forenduserdisplay>
      <c:name>Computer Hardware - NEC</c:name>
      <c:rcscode>B:1758</c:rcscode>
      <c:trbccode>5710601010</c:trbccode>
      <c:permid>4294951243</c:permid>
      <c:relevance>0.800</c:relevance>
   </rdf:Description>
```

[Industry Tag Attributes](#)

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e/Industry/3">
      <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/Industry"/>
      <c:docId rdf:resource="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-0103f889be3e"/>
      <c:forenduserdisplay>false</c:forenduserdisplay>
      <c:name>Phones &amp; Handheld Devices - NEC</c:name>
      <c:rcscode>B:1768</c:rcscode>
      <c:trbccode>5710602010</c:trbccode>
      <c:permid>4294951233</c:permid>
      <c:relevance>0.200</c:relevance>
   </rdf:Description>
```

[Industry Tag Attributes](#)

1.2.3   Additional Examples

For additional examples, and detailed information on parsing and interpreting the Open Calais response, see RDF Response Format, JSON Response Format, and N3 Response Format.

Note: It is important to note that most attributes are optional; a tag can be extracted with some but not all of its attributes.

1.3   Open Calais Premium

The Open Calais free service supports up to 5000 requests per day, and provides an <u>extensive set of semantic metadata tags</u>. If you are interested in processing a higher volume of data, or in the functionality provided by premium metadata types, please contact us at <u>questions@opencalais.com</u> for further information.

Existing Premium Customers, please contact us at <u>questions@opencalais.com</u>. We will be happy to assign a technical account manager to assist you with finding the best methods and tools for incorporating Open Calais metadata into your infrastructure and application.

Chapter 2   Forming the API Call

2.1   Resource URLs

Calls to tag content are made via a simple HTTP REST interface.

The Open Calais tagging method is located at POST https://api.thomsonreuters.com/permid/calais.

2.2   Request Authentication

In order to post a request, you are required to register and obtain an API access token at www.opencalais.com. If you are a registered user of the Thomson Reuters Open PermID services, you can use the same token for Calais Tagging. If you aren't familiar with our https://permid.org/ services, please check it out!

2.3   Input Content Type

Open Calais currently supports the following input content types: text/html, text/xml, and text/raw. In addition, Thomson Reuters Intelligent Tagging (TRIT) supports the application/pdf content type.

2.4   Input Size

The maximum input size is limited to 100KB (not characters, KB) per request.

For premium users, larger input files can be supported. Please contact us at questions@opencalais.com for further information.

A submission that exceeds the input size limit is not processed, and an error message is returned.

The input size limitation applies to the entire document, including the body and xml tags, but excluding the HTTP headers.

The input size limitation applies to all input file types (raw text, xml, html, pdf). (The pdf input file type is supported for TRIT users only).

Note: The size limitation defines the maximum file size that the system can process. However, processing time depends on the complexity of the text within the file, and a timeout error may be generated if a file is too complex (contains too many entities and relations) to be processed within the time limit.

2.5   Input Language

Open Calais currently supports English, French, and Spanish. For a full inventory of supported metadata for each language, see Open Calais Semantic Metadata Tags (English), French Semantic Metadata Tags, Spanish Semantic Metadata Tags.

Open Calais determines the language of the submitted content automatically, and invokes the appropriate extraction module for extracting entities and relations from the text. Open Calais may be unable to determine the language properly if the submitted content is too short.

You can use the x-calais-language request header to override the automatic language detection functionality.

2.6 Request Headers

The input content sent to Open Calais is accompanied by a set of parameters specified in key-value pairs as HTTP headers of the request. The parameters must be sent as US-ASCII characters. Header names and values are not case sensitive.

In case an HTTP header contains a non-US-ASCII character, the client application must encode it before sending this header to Open Calais.

The request is an HTTP Post with the following query parameters:

- Content-Type (**MANDATORY**): Indicates the input mime type.

- omitOutputtingOriginalText: Excludes the original text from the output. Highly recommended for large input files.

- outputFormat: Defines the output format.

- x-ag-access-token (**MANDATORY**): The value of this header is your license key.

- x-calais-contentClass: Specifies the genre of the input document. Highly recommended for optimal extraction when input files are research reports in PDF format, or news stories.

- x-calais-language: Indicates the language of the input text.

- x-calais-selectiveTags: Lets you specify a custom set of metadata tag types to be included in the output.

2.6.1 Content-Type (Mandatory)

Content-Type	
Description	Indicates the input content type (mime type). Open Calais processes the input documents according to the value of this parameter for optimal metadata extraction.
Values	• text/html: Use this value when submitting web pages. • text/xml: Use this value when submitting XML content. • text/raw: Use this value when submitting clean, unformatted text. • application/pdf: Use this value when submitting PDF files as binary streams. This value is available to Premium (Thomson Reuters Intelligent Tagging) users. Please make sure that your PDF files contain text objects; Open Calais does not extract text from images in PDF files.
Default Value	None
Remarks	• This is a mandatory parameter. • We recommend that before submission, you remove from the input document any redundant or irrelevant text (such as ads, disclaimers, repeated generic text such as "contact customer support for further advice…," trademarks, etc.). • Text content should be UTF-8 encoded; otherwise, specify charset, e.g. text/xml; charset=utf-8. • Note that if your text includes accented characters, for example, "Ségolène Royal," and you do not set encoding to UTF-8, the Open Calais output strips these characters, trashing the original text. • Open Calais expects the url-encoded arguments to be encoded using UTF-8. HttpClient defaults to another encoding, so you must instruct it to use UTF-8 for proper url-encoding of your arguments. • For binary documents (e.g. PDF) the http body should include the binary stream. • To optimize tagging of research reports (in PDF format only), make sure to also define the x-calais-contentClass header for best results.

2.6.2 omitOutputtingOriginalText

omitOutputtingOriginalText	
Description	Use this parameter to exclude the submitted text from the output, thus reducing the size of the response.
Values	• true • false
Default Value	false (The original text is included in the output.)
Remarks	By default, Open Calais returns the submitted text. Set this parameter to **true** to exclude the original text from the output. It is highly recommended to use this header for large input files

2.6.3   outputFormat

outputFormat	
Description	Defines the output response format (mime type).
Values	• xml/rdf • application/json • text/n3
Default Value	xml/rdf

2.6.4   x-ag-access-token

x-ag-access-token	
Description	The license key (token) which grants you access to Open Calais and defines your submission capacity rights.
Values	Your license key. If you do not already have a license key, you can sign up for one at www.opencalais.com.
Default Value	None
Remarks	• This is a mandatory parameter. • You should note the allowed submission size and rate for the token and adhere to them in your code; if you exceed these limits, your submissions can be blocked automatically for a certain period of time.

2.6.5   x-calais-contentClass

x-calais-contentClass	
Description	Lets you specify the genre of the input files, to optimize extraction.
Values	• news – Define this value when input files are news stories. • research – Define this value when input files are research reports in PDF format.
Default Value	None
Remarks	For best quality output, it is highly recommended to use this header when the input files are research reports in PDF format, or news stories.

2.6.6   x-calais-language

x-calais-language	
Description	Specifies the language of the input text. You can use this parameter to override the automatic language detection functionality.
Values	The full name of the language, in English: • English • French • Spanish
Default Value	None
Remarks	If you know the language you are submitting, it is recommended to pass this value in the API. It is especially recommended if your texts are short or have many non-letter symbols in them, which reduces the accuracy of automatic language identification.

2.6.7   x-calais-selectiveTags

x-calais-selectiveTags	
Description	Limits the output to selected metadata tag types. You can select multiple values to define a specific set of output metadata tags.
Values	Valid values: **additionalcontactdetails**, **company**, **country**, **deal,company**, **industry, person**, **socialtags**, **topic**, **topic-sca, topic-selfservice** **additionalcontactdetails** – the output file includes the following: • em/e/EmailAddress • em/e/IndustryTerm • em/e/FaxNumber • em/e/PhoneNumber • Related InstanceInfo tags • Related RelevanceInfo tags • All extracted relations, including em/r/PersonAttributes, em/r/PersonRelation, and em/r/PersonTravel tags. This header value does not trim relations tags (em/r/ tags) from the output. • Any additional entities that the extracted relations link to. For example, if an em/r/PersonAttributes tag is extracted, and it defines a *person* attribute which is a link to a person, then the relevant em/e/Person tag is also included in the output. **company** – the output file includes the following: • All extracted company tags (em/e/Company, er/Company, er/TopmostPublicParentCompany). • The related InstanceInfo, Confidence, and RelevanceInfo tags. **country** – the output file includes the following: • All extracted country tags (em/e/Country, er/Geo/Country) • The related InstanceInfo and RelevanceInfo tags **deal,company** – the output file includes the following: • All extracted deal tags (em/r/deal, er/deal) and related InstanceInfo and Confidence tags. • All extracted company tags (em/e/Company, er/Company, er/TopmostPublicParentCompany) and the related InstanceInfo, Confidence, and RelevanceInfo tags. • All extracted relations. This header value does not trim relations tags (em/r/ tags) from the output. • Any additional entities that the extracted relations link to. For example, if an em/r/Quotation tag is extracted, and the *speaker* attribute value is a link to a person, then the linked em/e/Person tag is included in the output as well. **industry**– the output file is limited to Industry tags. **person** – the output file includes the following: • All extracted person tags (em/e/Person, er/Person). • The related InstanceInfo, Confidence, and RelevanceInfo tags. **socialtags** – the output is limited to SocialTag tags. **topic** – the output is limited to DocCat (topic) tags for RCS topics.
Default Value	None
Remarks	You can define multiple values.

2.7   A Sample Request File

```
HTTPS REST7
POST https://api.thomsonreuters.com/permid/calais
```

Parameters

```
Content-Type: text/xml
outputFormat:xml/rdf
x-ag-access-token: (your authorized token)
```

Body

```
<Document>
    <Title>GoPro launches $800 million offering, CEO to sell some shares</Title>
    <Body>
        Wearable camera maker GoPro Inc's chief executive, Nicholas Woodman, plans to sell a portion of his
        stake as part of an $800 million offering of the company's shares. The offering of Class A common shares
        is expected to start in the next couple of weeks and close by November, a company spokesman told
        Reuters. GoPro's shares fell as much as 5.1 percent before easing back a little to trade down 3.8
        percent at $76.04 on Monday. The lock-up period on the stock, which listed in June, expires on Dec. 22,
        allowing employees and early investors to sell shares of the company. Typically, on the day a lock-up
        expires, prices tend to fall as a large number of shares become available for trading. The company said
        the offering was expected to soften the blow of the lock-up expiration on the share price.  (Additional
        reporting by Arathy Nair; Editing by Saumyadeb Chakrabarty)
    </Body>
</Document>
```

For a detailed explanation of Open Calais output see The API Response.

2.8   Security

We respect and guard the privacy of your information. We implement technological safeguards to prevent unauthorized access to your data and we have made a point of implementing a processing workflow that negates any requirement for us to store your data or metadata.

Open Calais operates entirely over HTTPS in order to secure traffic to and from Open Calais. The secure connection (HTTPS) ensures that the information you send for processing remains private.

Your data simply passes through our processing engine. Open Calais processes your data and returns metadata. Open Calais does not store your data; Open Calais does not store your metadata.

If you have any questions or concerns, please contact us.

Chapter 3   The API Response

This chapter describes the following:

REST Interface

OWL Schema

RDF Response Format

JSON Response Format

N3 Response Format

Error Messages

3.1   REST Interface

The Open Calais web service follows standard REST web service guidelines. This mainly involves responding to the client with standard HTTP error codes.

All successful HTTP transactions return the HTTP status of **200 OK**.

Appropriate load/processing errors such as the server is under load, the language or format of document is not supported, etc., are returned in the body of the HTTP response.

Internal Open Calais exceptions due to unknown reasons result in HTTP responses with the exception/message in the response body and HTTP error code 500.

In addition, the client must take into account issues usually associated with HTTP transactions, such as networking, transport protocol and system issues. An example would be any erroneous HTTP status-code (other than 200), tcp-timeout, tcp-connection-closed, etc., which was reported by the low-level networking layers.

3.2 OWL Schema

Open Calais OWL is an ontology which describes the data structure of the Open Calais output. The OWL describes the metadata types that can be output by Open Calais, their possible attributes and the relevant constraints. The OWL also describes relations such as inheritance and directional referencing between metadata elements.

Note: It is important to note that most attributes are optional; a tag can be extracted with some but not all of its attributes.

The OWL can be used as a reference for understanding the range of Calais metadata elements and their relationships to one another. You can download the current OWL schema from the API page of www.opencalais.com.

3.3   RDF Response Format

This section provides an RDF output sample with explanation as a general guideline to parsing and interpreting the Open Calais response. The Open Calais response is based on an OWL schema. You can download the current OWL schema from the API page of www.opencalais.com.

Note: It is important to note that most attributes are optional; a tag can be extracted with some but not all of its attributes.

The output file is comprised of the following main sections:

- General Document Information

- Aboutness Tags that describe the document as a whole.

- Named Entity and Relationship Recognition Tags that describe the individual text strings contained within the document.

Note: Open Calais also supports integration with third party NLPs. See Third Party NLP Onboarding.

3.3.1 General Document Information

[Summary of Extracted Relations and Entities](#)

[DocInfo](#)

[DocInfoMeta](#)

[ComponentVersions](#)

[DefaultLangID](#)

3.3.1.1   Summary of Extracted Relations and Entities

The output includes a summary of all relations and entities extracted from the text. Extracted entities are displayed alphabetically and grouped by type.

Take a look at the output example, below, and note the following:

- The extracted relations are CompanyExpansion, CompanyLocation, CompanyTechnology, PersonLocation, Quotation.

- The extracted entity types are City, Company, Country, Editor, Facility, IndustryTerm, Journalist, Position, Product, ProvinceOrState, PublishedMedium, Technology.

- The extracted City entities are Cupertino and London; the extracted Company entities are Apple, Audi, BMW, Daimler, etc.; the extracted Country entity is Austria; etc.

```
<!--Relations: CompanyExpansion, CompanyLocation, CompanyTechnology, PersonLocation, Quotation,
City: Cupertino, London
Company: Apple, Audi, BMW, Daimler, Google Inc, Magna International, Reuters, Tesla Motors Inc, Volkswagen, Wall
        Street Journal, the Financial Times
Country: Austria
Editor: Hugh Lawson, Noah Barkin
Facility: Apple's headquarters
IndustryTerm: Car technology, actual car, automated driving technology, automotive software, automotive
        technology, car, car recharging services, carmakers and automotive suppliers, driver assistance systems,
        electric and connected-car technologies, electric car, mobile and electronic devices, prototype self-
        driving car, self-driving electric car, software game, studies self-driving car
Journalist: Eric Auchard, Hugh Lawson
Position: driver, spokesman, spokeswoman
Product: iPad, iPhone
ProvinceOrState: California
PublishedMedium: The Wall Street Journal, the Financial Times
Technology: Car technology, automated driving technology, electric and connected-car technologies, mobile
        devices
-->
```

3.3.1.2   DocInfo

The DocInfo node presents the original input text (inside the body tag).

Note the dochash, which appears in many of the tags in the output RDF; it is the unique ID of the containing document.

```
<rdf:Description c:calaisRequestID="b93c8375-d19d-d574-14bf-8f4b13366849"
    c:id="http://id.opencalais.com/qXNdmiJnB1x3doCaXtkMKA" c:ontology="http://och1-
    1b/owlschema/8.2/onecalais.owl.allmetadata.xml" rdf:about="http://d.opencalais.com/dochash-1/7586b818-
    40af-3d55-ac16-bf3520cddfa6">
    <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/DocInfo"/>
    <c:document>
        <![CDATA[<Document>
<Title>Apple studies self-driving car, auto industry source says</Title>
<Date>2015-02-14</Date>
<Body>
(Reuters) - Technology giant Apple (AAPL.0) is looking beyond mobile devices to learn how to make a self-
    driving electric car, and is talking to experts at carmakers and automotive suppliers, a senior auto
    industry source familiar with the discussions said on Saturday. The Cupertino, California-based maker of
    phones, computers and, soon, watches is exploring how to make an entire vehicle, not just designing
    automotive software or individual components, the auto industry source said. "They don't appear to want
    a lot of help from carmakers," said the source, who declined to be named. Apple is gathering advice on
    parts and production methods, focusing on electric and connected-car technologies, while studying the
    potential for automated driving, the source said. "Fully automated driving is an evolution. Carmakers
    will slowly build the market for autonomous cars by first releasing connected and partially automated
    cars," the auto industry source said. "Apple is interested in all the potential ways you can evolve the
    car; that includes autonomous driving." Whether it will build and release an electric car or a more
    evolved autonomous vehicle remains to be seen, the source said. But clearly Apple has sharply raised its
    ambitions in automotive technology.

    Car technology has become a prime area of interest for Silicon Valley companies ranging from Google Inc
    (GOOGL.0), which has built a prototype self-driving car, to electric car-maker Tesla Motors Inc
    (TSLA.0). An Apple spokesman in London on Saturday declined to comment on "rumors or speculation".
    Trying to build an actual car would mark a dramatic shift for the maker of the iPhone and iPad. Apple
    often researches projects which are then discarded, but has so far mainly stuck to its core expertise in
    mobile and electronic devices.

    The Wall Street Journal reported on Friday that Apple had set up a secret lab working on the creation of
    an Apple-branded electric car, citing people familiar with the matter. The lab was set up late last
    year, soon after Apple revealed its forthcoming smart watch and latest iPhones, the Financial Times
    said. The Journal said that the project, code-named "Titan", employed several hundred people working a
    few miles from Apple's headquarters in Cupertino. Apple executives met with contract manufacturers
    including Magna Steyr in Austria, a unit of Magna International (MG.TO), the Journal said. A Magna
    spokeswoman declined to comment.

    THE PATH TO SELF-DRIVING CARS
    Autonomous driving is likely to emerge progressively as driver assistance systems become more
    sophisticated. Already, carmakers such as Daimler (DAIGn.DE), BMW (BMWG.DE) and Volkswagen's (VOWG_p.DE)
    Audi (NSUG.DE) have revealed cars that can travel long distances without human intervention. Analysts at
    Exane BNP Paribas have said they see a $25 billion market for automated driving technology by 2020, with
    vehicle intelligence becoming "the key differentiating factor". But the brokerage does not expect fully
    automated cars to hit the road until 2025 or 2030, in part due to regulatory hurdles.

    Short of building entire cars, there is money to be made from the software to run a self-driving
    vehicle, as well as the services associated with autonomous driving, such as mapping, car-sharing and
    car recharging services, the auto source said. "It's a software game. It's all about autonomous
    driving," the industry source said.

    (Additional reporting by Eric Auchard in Frankfurt; Editing by Noah Barkin/Hugh Lawson)
    </Body>
</Document>]]></c:document>
    <c:docTitle>Apple studies self-driving car, auto industry source says</c:docTitle>
    <c:docDate>2015-02-14 00:00:00</c:docDate>
</rdf:Description>
```

3.3.1.3   DocInfoMeta

The DocInfoMeta node presents processing information such as the submission date and time, and the identity of the submitter.

```
<rdf:Description c:contentType="text/xml" c:language="English" c:processingVer="AllMetadata"
    c:serverVersion="OneCalais_8.2-RELEASE:360" c:stagsVer="OneCalais_8.2-RELEASE-b6-2015-02-14_17:44:23"
    c:submissionDate="2015-03-10 14:14:50.994" rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-
    3d55-ac16-bf3520cddfa6/meta">
    <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/DocInfoMeta"/>
    <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
    <c:submitterCode>d9d88048-1255-f96b-87c2-22d93db1bd23</c:submitterCode>
    <c:signature>digestalg-
    1|6u2soqBju3OYouXxdALFwyPy7vQ=|bWW7KmieILQ6W/nmZ4uh8v6r8d4xFIjW5cWy/jgBYzsNFUzMKMRvxA==</c:signature>
</rdf:Description>
```

3.3.1.4 Component Versions

The ComponentVersions node specifies the component versions used to process the input file. This information is primarily for the use of the Text Metadata Server (TMS) group at Thomson Reuters, in case of a processing problem.

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
    bf3520cddfa6/ComponentVersions">
    <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/ComponentVersions"/>
    <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
    <c:version>Deals Index: 201503080600: 201503080600</c:version>
    <c:version>OA Index: 201503080155: 201503080155</c:version>
    <c:version>NextTags: OneCalais_8.2-RELEASE: 108</c:version>
    <c:version>SpanishIM: OneCalais_8.2-RELEASE: 195</c:version>
    <c:version>config-sca-DataPackage: 34: 34</c:version>
    <c:version>SECHeaderMetadataIM: OneCalais_8.2-RELEASE: 195</c:version>
    <c:version>com.clearforest.infoext.dial4j.plugins-basistechconfig: OneCalais_8.2-RELEASE: 222</c:version>
    <c:version>People Index: 201503071945: 201503071945</c:version>
    <c:version>Collector: OneCalais_8.2-RELEASE: 108</c:version>
    <c:version>Dial4J: OneCalais_8.2-RELEASE: 195</c:version>
    <c:version>AutocoderRuntimeIM: OneCalais_8.2-RELEASE: 195</c:version>
    <c:version>OA Override: 251: 251</c:version>
    <c:version>People Override: 247: 247</c:version>
    <c:version>BrokerResearchIM: OneCalais_8.2-RELEASE: 195</c:version>
    <c:version>config-refineries: 247: 247</c:version>
    <c:version>config-cse: 247: 247</c:version>
    <c:version>OneCalaisIM: OneCalais_8.2-RELEASE: 195</c:version>
    <c:version>config-vessels: 247: 247</c:version>
    <c:version>OneCalais: OneCalais_8.2-RELEASE: 360</c:version>
    <c:version>Housekeeper: OneCalais_8.2-RELEASE: 108</c:version>
    <c:version>WatchDog: OneCalais_8.2-RELEASE: 108</c:version>
    <c:version>SocialTags Index: 201503080540: 201503080540</c:version>
    <c:version>BlackList: 247: 247</c:version>
    <c:version>FrenchIM: OneCalais_8.2-RELEASE: 195</c:version>
    <c:version>config-physicalAssets-ports: 247: 247</c:version>
    <c:version>config-drugs: 247: 247</c:version>
</rdf:Description>
```

3.3.1.5   DefaultLangID

This tag indicates the input text language, specified by the <u>x-calais-language</u> request header, or identified by Open Calais.

Based on the input text language, Open Calais invokes the appropriate metadata extraction module.

In the following example, the indicated input text language is English.

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
    bf3520cddfa6/lid/DefaultLangId">
    <rdf:type rdf:resource="http://s.opencalais.com/1/type/lid/DefaultLangId"/>
    <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
    <c:lang rdf:resource="http://d.opencalais.com/lid/DefaultLangId/English"/>
    <c:forenduserdisplay>false</c:forenduserdisplay>
    <c:permid>505062</c:permid>
</rdf:Description>
```

Note the *forenduserdisplay* attribute. This is our recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).

3.3.2   Aboutness Tags

Aboutness tags describe the piece of content as a whole.

The Open Calais response may include any of the following metadata tags:

- SocialTag
- DocCat (Topic tag)
- Industry

3.3.2.1   SocialTag

Social tags attempt to classify the document as a whole, based on Wikipedia folksonomy.

Examples of social tags, extracted by Open Calais from the article about Apple developing a self-driving car:

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
    bf3520cddfa6/SocialTag/1">
    <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/SocialTag"/>
    <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
    <c:socialtag rdf:resource="http://d.opencalais.com/genericHasher-1/1205cb52-d703-34d2-83b2-
    a09d4d47575c"/>
    <c:forenduserdisplay>true</c:forenduserdisplay>
    <c:name>Apple Inc.</c:name>
    <c:importance>1</c:importance>
    <c:originalValue>Apple Inc.</c:originalValue>
</rdf:Description>
```

SocialTag Attributes

Note the *importance* attribute, which indicates how centric the topic named by the social tag is to the document as a whole. The importance attribute value can be 1 (very centric), 2 (somewhat centric), or 3 (less centric).

The *docID* (dochash) is the unique ID of the containing document.

The *rdf:Description* is the tag's unique ID within the containing document. The dochash component of the rdf:Description is what associates this tag with this document.

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
    bf3520cddfa6/SocialTag/2">
    <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/SocialTag"/>
    <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
    <c:socialtag rdf:resource="http://d.opencalais.com/genericHasher-1/ccc60460-211d-3b02-b13e-
    11fba4449fbd"/>
    <c:forenduserdisplay>true</c:forenduserdisplay>
    <c:name>Autonomous car</c:name>
    <c:importance>1</c:importance>
    <c:originalValue>Autonomous car</c:originalValue>
</rdf:Description>
```

SocialTag Attributes

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
    bf3520cddfa6/SocialTag/4">
    <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/SocialTag"/>
    <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
    <c:socialtag rdf:resource="http://d.opencalais.com/genericHasher-1/083d56d1-2fed-3063-b59a-
    963d4fdaee36"/>
    <c:forenduserdisplay>true</c:forenduserdisplay>
    <c:name>Electric car</c:name>
    <c:importance>2</c:importance>
    <c:originalValue>Electric car</c:originalValue>
</rdf:Description>
```

SocialTag Attributes

3.3.2.2 DocCat (Topic Tags)

Topic tags identify the topic or topics being discussed in the document. The reference list of topics is defined by the Thomson Reuters Coding Schema (TRCS) and/or by the International Press Telecommunications Council (IPTC) news taxonomy and/or by the Self Service Classification project taxonomy. See Supported Topics.

Note that access to TRCS is available upon subscription to Thomson Reuters Intelligent Tagging (TRIT). For more information please contact us at questions@opencalais.com.

Examples of topic tags, extracted by Open Calais from the article about Apple developing a self-driving car.

An IPTC taxonomy topic:

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6/cat/1">
    <rdf:type rdf:resource="http://s.opencalais.com/1/type/cat/DocCat"/>
    <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
    <c:forenduserdisplay>false</c:forenduserdisplay>
    <c:score>0.988</c:score>
    <c:name>Technology_Internet</c:name>
</rdf:Description>
```

Topic Tag (DocCat tag) attributes

Note the *score* attribute, which indicates the probability, on a scale of 0 to 1 that the topic is indeed discussed in the text, and also how centric the topic is to the text; the higher the value, the higher the probability.

A TRCS taxonomy topic:

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6/cat/2">
    <rdf:type rdf:resource="http://s.opencalais.com/1/type/cat/DocCat"/>
    <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
    <c:forenduserdisplay>false</c:forenduserdisplay>
    <c:rcscode>B:172</c:rcscode>
    <c:name>Software &amp; IT Services (TRBC)</c:name>
    <c:shortName>Software &amp; IT Services</c:shortName>
    <c:score>0.765</c:score>
</rdf:Description>
```

Topic Tag (DocCat tag) attributes

3.3.2.3   Industry Tags

Industry tags indicate the industries that are related to the companies mentioned in the input text. The reference list of industries is defined by the Thomson Reuters Business Classification (TRBC) taxonomy.

The following Industry tags were extracted by Open Calais from the story about Apple developing a self-driving car.

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
    bf3520cddfa6/Industry/5">
    <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/Industry"/>
    <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
    <c:forenduserdisplay>false</c:forenduserdisplay>
    <c:name>Automobiles &amp; Multi Utility Vehicles</c:name>
    <c:rcscode>B:1294</c:rcscode>
    <c:trbccode>5210101097</c:trbccode>
    <c:permid>4294951707</c:permid>
    <c:relevance>0.500</c:relevance>
</rdf:Description>
```

Industry Tag Attributes

Industry tags include a unique Thomson Reuters ID (the *permid* attribute value). The ID can be used to extract information about the industry from the Thomson Reuters dataset. The ID also supports linkage across documents processed by Open Calais.

Note the *relevance* attribute, which indicates how relevant the industry is to the story. Values range from 0 to 1; the higher the score, the higher the relevance.

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
    bf3520cddfa6/Industry/7">
    <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/Industry"/>
    <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
    <c:forenduserdisplay>false</c:forenduserdisplay>
    <c:name>Electrical (Alternative) Vehicles</c:name>
    <c:rcscode>B:1296</c:rcscode>
    <c:trbccode>5210101025</c:trbccode>
    <c:permid>4294951705</c:permid>
    <c:relevance>0.500</c:relevance>
</rdf:Description>
```

Industry Tag Attributes

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
    bf3520cddfa6/Industry/8">
    <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/Industry"/>
    <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
    <c:forenduserdisplay>false</c:forenduserdisplay>
    <c:name>Auto &amp; Truck Manufacturers - NEC</c:name>
    <c:rcscode>B:1292</c:rcscode>
    <c:trbccode>5210101029</c:trbccode>
    <c:permid>4294951709</c:permid>
    <c:relevance>0.500</c:relevance>
</rdf:Description>
```

[Industry Tag Attributes](#)

3.3.3 Named Entity and Relationship Recognition

Open Calais scans and analyzes the input text, searching for mentions of things like companies, people, deals, and geographical locations, based on a <u>list of predefined metadata types</u>. The resulting tags form the major part of the output response: <u>Instance tags</u>, <u>Entity Markup tags</u>, <u>Relevance tags</u>, <u>Confidence tags</u>, and <u>Disambiguation tags</u>.

In this section we'll highlight some of the metadata tags extracted by Open Calais from the article about Apple developing a self-driving car.

One of the Open Calais predefined metadata types is Company. The article mentions several companies. Let's take a look at the InstanceInfo tags generated by Open Calais for a few of the found mentions of the company, Apple:

"Apple studies self-driving car, auto industry source says"

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6/Instance/59">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/InstanceInfo"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
        <c:subject rdf:resource="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4"/>
        <!--Company: Apple; -->
        <c:detection>[&lt;Document&gt; &lt;Title&gt;]Apple[ studies self-driving car, auto industry
source]</c:detection>
        <c:prefix>&lt;Document&gt; &lt;Title&gt;</c:prefix>
        <c:exact>Apple</c:exact>
        <c:suffix> studies self-driving car, auto industry source</c:suffix>
        <c:offset>20</c:offset>
        <c:length>5</c:length>
    </rdf:Description>
```

<u>InstanceInfo Tag Attributes</u>

"Technology giant Apple (APPL.O) is looking beyond mobile devices to learn how to make a self-driving electric car, and is…"

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6/Instance/60">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/InstanceInfo"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
        <c:subject rdf:resource="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4"/>
        <!--Company: Apple; -->
        <c:detection>[&lt;Body&gt; (Reuters) - Technology giant ]Apple (AAPL.0)[ is looking beyond mobile
devices to learn how to]</c:detection>
        <c:prefix> &lt;Body&gt; (Reuters) - Technology giant </c:prefix>
        <c:exact>Apple (AAPL.0)</c:exact>
        <c:suffix> is looking beyond mobile devices to learn how to</c:suffix>
        <c:offset>151</c:offset>
        <c:length>14</c:length>
    </rdf:Description>
```

<u>InstanceInfo Tag Attributes</u>

"An Apple spokesman in London on Saturday declined to comment on 'rumors or speculation.' "

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6/Instance/63">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/InstanceInfo"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
        <c:subject rdf:resource="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4"/>
        <!--Company: Apple; -->
        <c:detection>[car-maker Tesla Motors Inc (TSLA.0). An ]Apple[ spokesman in London on Saturday declined
        to]</c:detection>
        <c:prefix>car-maker Tesla Motors Inc (TSLA.0). An </c:prefix>
        <c:exact>Apple</c:exact>
        <c:suffix> spokesman in London on Saturday declined to</c:suffix>
        <c:offset>1625</c:offset>
        <c:length>5</c:length>
    </rdf:Description>
```

<u>InstanceInfo Tag Attributes</u>

"The Journal said that the project, code-named 'Titan,' employed several hundred people working a few miles from Apple's headquarters in Cupertino."

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6/Instance/68">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/InstanceInfo"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
        <c:subject rdf:resource="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4"/>
        <!--Company: Apple; -->
        <c:detection>[several hundred people working a few miles from ]Apple['s headquarters in Cupertino. Apple
        executives]</c:detection>
        <c:prefix>several hundred people working a few miles from </c:prefix>
        <c:exact>Apple</c:exact>
        <c:suffix>'s headquarters in Cupertino. Apple executives</c:suffix>
        <c:offset>2387</c:offset>
        <c:length>5</c:length>
    </rdf:Description>
```

<u>InstanceInfo Tag Attributes</u>

Each group of one or more instances deemed to refer to a unique thing is expressed as an Entity Markup tag. The following example is the Entity Markup (em/e/company) tag for the company entity, Apple.

The hash tag (comphash) is the unique ID assigned by Open Calais to this extracted entity. This hash tag also appears in the associated instance tags.

```
<rdf:Description rdf:about="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/em/e/Company"/>
        <c:forenduserdisplay>true</c:forenduserdisplay>
        <c:name>Apple</c:name>
        <c:nationality>American</c:nationality>
        <c:confidencelevel>0.996</c:confidencelevel>
    </rdf:Description>
```

<u>em/e/Company Tag Attributes</u>

The high *confidencelevel* value indicates confidence that the extracted company, Apple, is indeed a company.

Some tag types that include the *confidencelevel* attribute also have an associated Confidence tag:

```
<rdf:Description rdf:about="http://d.opencalais.com/conf/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/Confidence"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
        <!--Apple-->
        <c:subject rdf:resource="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4"/>
        <c:statisticalfeature>0.997</c:statisticalfeature>
        <c:dblookup>0.0</c:dblookup>
        <c:resolution>0.9928677</c:resolution>
        <c:aggregate>0.996</c:aggregate>
    </rdf:Description>
```

<u>Confidence Tag Attributes</u>

Again, note that the entity markup tag and all of its associated tags display the same comphash.

The *aggregate* attribute value is the confidence score.

The RelevanceInfo tag associated with the company entity, Apple, indicates that Apple is highly relevant to the story:

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
        bf3520cddfa6/Relevance/47">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/RelevanceInfo"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
        <c:subject rdf:resource="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4"/>
        <c:relevance>0.8</c:relevance>
        <c:relevancecont>0.72</c:relevance>
    </rdf:Description>
```

<u>RelevanceInfo Tag Attributes</u>

Compare the RelevanceInfo tag for the extracted company, Apple, with the Relevance tag associated with the extracted company, Daimler, which was mentioned in the document, along with several other companies, as having "…revealed cars that can travel long distances without human intervention…:"

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
        bf3520cddfa6/Relevance/36">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/RelevanceInfo"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
        <c:subject rdf:resource="http://d.opencalais.com/comphash-1/ebd24b16-1574-33c0-9b69-30c054337d44"/>
        <c:relevance>0.5</c:relevance>
        <c:relevancecont>0.568</c:relevance>
    </rdf:Description>
```

<u>RelevanceInfo Tag Attributes</u>

Note that the comphash assigned to Daimler is different from the comphash assigned to Apple.

Open Calais succeeded in mapping the extracted company entity, Apple, to the corresponding company and unique ID in the Thomson Reuters dataset, resulting in the following Disambiguation (er/company) tag:

```
<rdf:Description rdf:about="http://d.opencalais.com/er/company/ralg-oa/4295905573">
      <rdf:type rdf:resource="http://s.opencalais.com/1/type/er/Company"/>
      <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
      <c:permid>4295905573</c:permid>
      <c:score>0.9928677</c:score>
      <!--Apple-->
      <c:subject rdf:resource="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4"/>
      <c:legacyid rdf:resource="http://d.opencalais.com/er/company/ralg-tr1r/23d07771-c50b-315b-8050-
      3cdaf47ac0d0"/>
      <c:name>Apple Inc</c:name>
      <c:commonname>Apple</c:commonname>
      <c:ticker>AAPL</c:ticker>
      <c:primaryric>AAPL.OQ</c:primaryric>
      <c:ispublic>true</c:ispublic>
      <c:openpermid rdf:resource="https://permid.org/1-4295905573"/>
   </rdf:Description>
```

er/company Tag Attributes

The comphash assigned by Open Calais identifies and links all the associated tags within the output document (local disambiguation).

The successful mapping to the corresponding Thomson Reuters entity and permid enables the Open Calais entity, in this case the company, Apple, Inc., to be unambiguously identified (and thus linked) across all documents processed by Open Calais.

Note: The **openpermid** attribute gives you direct access to high quality, curated Thomson Reuters company data. The attribute value is a direct link to the relevant company page on the Open PermID website (https://permid.org).

The following instance tag was generated when Open Calais found a mention of a company expansion:

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6/Instance/19">
      <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/InstanceInfo"/>
      <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
      <c:subject rdf:resource="http://d.opencalais.com/genericHasher-1/4c3556cd-1c91-363f-b4b2-a9d1b372aa85"/>
      <!--CompanyExpansion: company: Apple; expansiontype: New Unit; status: known; -->
      <c:detection>[ The Wall Street Journal reported on Friday that ]Apple had set up a secret lab[ working
      on the creation of an Apple-branded]</c:detection>
      <c:prefix> The Wall Street Journal reported on Friday that </c:prefix>
      <c:exact>Apple had set up a secret lab</c:exact>
      <c:suffix> working on the creation of an Apple-branded</c:suffix>
      <c:offset>2003</c:offset>
      <c:length>29</c:length>
   </rdf:Description>
```

InstanceInfo Tag Attributes

The resulting Entity Markup (em/r/CompanyExpansion) tag (the extracted relation):

```
<rdf:Description rdf:about="http://d.opencalais.com/genericHasher-1/4c3556cd-1c91-363f-b4b2-a9d1b372aa85">
      <rdf:type rdf:resource="http://s.opencalais.com/1/type/em/r/CompanyExpansion"/>
      <c:forenduserdisplay>false</c:forenduserdisplay>
      <!--Apple-->
      <c:company rdf:resource="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4"/>
      <c:expansiontype>New Unit</c:expansiontype>
      <c:status>known</c:status>
</rdf:Description>
```

[em/r/CompanyExpansion Tag Attributes](#)

Note that the genericHasher in the *rdf:Description* links this Entity Markup tag to the associated InstanceInfo tag.

Also note that the *company* attribute value is a reference to the extracted company entity, Apple.

The following instance tag was generated when Open Calais found a text string that it identified as an indication of a relationship between a company and a technology (i.e. a mention of the type CompanyTechnology):

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6/Instance/27">
      <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/InstanceInfo"/>
      <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
      <c:subject rdf:resource="http://d.opencalais.com/genericHasher-1/55d6ef1d-6140-3150-930c-0d5c2fbd0107"/>
      <!--CompanyTechnology: company: Apple; technology: mobile devices; -->
      <c:detection>[&lt;Date&gt;2015-02-14&lt;/Date&gt; &lt;Body&gt; (Reuters) - ]Technology giant Apple
      (AAPL.O) is looking beyond mobile devices[ to learn how to make a self-driving electric]</c:detection>
      <c:prefix> &lt;Date&gt;2015-02-14&lt;/Date&gt; &lt;Body&gt; (Reuters) - </c:prefix>
      <c:exact>Technology giant Apple (AAPL.O) is looking beyond mobile devices</c:exact>
      <c:suffix> to learn how to make a self-driving electric</c:suffix>
      <c:offset>134</c:offset>
      <c:length>64</c:length>
   </rdf:Description>
```

[InstanceInfo Tag Attributes](#)

The resulting Entity Markup (em/r/CompanyTechnology) tag:

```
<rdf:Description rdf:about="http://d.opencalais.com/genericHasher-1/55d6ef1d-6140-3150-930c-0d5c2fbd0107">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/em/r/CompanyTechnology"/>
        <c:forenduserdisplay>false</c:forenduserdisplay>
        <!--Apple-->
        <c:company rdf:resource="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4"/>
        <!--mobile devices-->
        <c:technology rdf:resource="http://d.opencalais.com/genericHasher-1/38c58bd0-2536-3f03-bfa0-
        be1867f6fce8"/>
    </rdf:Description>
```

<u>em/r/CompanyTechnology Tag Attributes</u>

Again, note the genericHasher in the *rdf:Description* that links this Entity Markup tag to the associated InstanceInfo tag, and the *company* attribute value, the comphash which identifies the company, Apple Inc. The *technology* attribute value references the relevant em/e/technology entity, not illustrated here.

The following instance tag was generated when Open Calais found a text string that it identified as a reference to a City:

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6/Instance/36">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/InstanceInfo"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
        <c:subject rdf:resource="http://d.opencalais.com/genericHasher-1/752be8ce-c588-3bbe-8526-af3b60708561"/>
        <!--City: Cupertino; -->
        <c:detection>[with the discussions said on Saturday. The ]Cupertino[, California-based maker of phones,
        computers]</c:detection>
        <c:prefix>with the discussions said on Saturday. The </c:prefix>
        <c:exact>Cupertino</c:exact>
        <c:suffix>, California-based maker of phones, computers</c:suffix>
        <c:offset>399</c:offset>
        <c:length>9</c:length>
    </rdf:Description>
```

<u>InstanceInfo Tag Attributes</u>

The resulting Entity Markup (em/e/city) tag:

```
<rdf:Description rdf:about="http://d.opencalais.com/genericHasher-1/752be8ce-c588-3bbe-8526-af3b60708561">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/em/e/City"/>
        <c:forenduserdisplay>false</c:forenduserdisplay>
        <c:name>Cupertino</c:name>
    </rdf:Description>
```

<u>em/e/City Tag Attributes</u>

The associated Disambiguation (er/Geo/City) tag was generated when Open Calais succeeded in mapping the extracted city entity, Cupertino, to the corresponding city in the Thomson Reuters dataset:

```
<rdf:Description rdf:about="http://d.opencalais.com/er/geo/city/ralg-geo1/4edd5509-ee4d-deba-738c-fc3900017f64">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/er/Geo/City"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
        <c:name>Cupertino, California, United States</c:name>
        <c:shortname>Cupertino</c:shortname>
        <c:latitude>37.3231</c:latitude>
        <c:longitude>-122.0311</c:longitude>
        <c:containedbystate>California</c:containedbystate>
        <c:containedbycountry>United States</c:containedbycountry>
        <!--Cupertino-->
        <c:subject rdf:resource="http://d.opencalais.com/genericHasher-1/752be8ce-c588-3bbe-8526-af3b60708561"/>
    </rdf:Description>
```

[er/Geo/City Tag Attributes](#)

The following InstanceInfo tag was generated when Open Calais found a text string that it identified as an indication of a relationship between a company and a location (i.e. a mention, of the type CompanyLocation):

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6/Instance/33">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/InstanceInfo"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
        <c:subject rdf:resource="http://d.opencalais.com/genericHasher-1/f64cef62-b4a7-39f1-b92a-ef8834053ff1"/>
        <!--CompanyLocation: company: Apple; companylocationtype: N/A; city: Cupertino; provinceorstate:
        California; -->
        <c:detection>[with the discussions said on Saturday. ]The Cupertino, California-based maker[ of phones,
        computers and, soon, watches is]</c:detection>
        <c:prefix>with the discussions said on Saturday. </c:prefix>
        <c:exact>The Cupertino, California-based maker</c:exact>
        <c:suffix> of phones, computers and, soon, watches is</c:suffix>
        <c:offset>395</c:offset>
        <c:length>37</c:length>
    </rdf:Description>
```

[InstanceInfo Tag Attributes](#)

The resulting Entity Markup (em/r/CompanyLocation) tag:

```
<rdf:Description rdf:about="http://d.opencalais.com/genericHasher-1/f64cef62-b4a7-39f1-b92a-ef8834053ff1">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/em/r/CompanyLocation"/>
        <c:forenduserdisplay>true</c:forenduserdisplay>
        <!--Apple-->
        <c:company rdf:resource="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4"/>
        <c:companylocationtype>N/A</c:companylocationtype>

        <!--Cupertino-->
        <c:city rdf:resource="http://d.opencalais.com/genericHasher-1/752be8ce-c588-3bbe-8526-af3b60708561"/>
        <!--California-->
        <c:provinceorstate rdf:resource="http://d.opencalais.com/genericHasher-1/9679b237-33e8-3478-ba13-
        d9af3c4b943e"/>
    </rdf:Description>
```

em/r/CompanyLocation Tag Attributes

The genericHasher in the *rdf:Description* links this tag to the related InstanceInfo tag; the *company* attribute value is the comphash that identifies Apple, Inc.; the *city* attribute value is the genericHasher that identifies the city, Cupertino.

3.4 JSON Response Format

This section provides a JSON output sample with explanation as a general guideline to parsing and interpreting the Open Calais response.

Note: It is important to note that most attributes are optional; a tag can be extracted with some but not all of its attributes.

The output file is comprised of the following main sections:

- General Document Information

- Aboutness Tags that relate to the document as a whole.

- Named Entity and Relationship Recognition tags that describe the text strings contained within the document.

3.4.1 General Document Information

Info

Meta

Component Versions

DefaultLangID

3.4.1.1   Info

The Info node presents the original input text (inside the body tag).

Note the dochash, which appears in many of the tags in the output file; it is the unique ID of the containing document.

```
"info": {
    "calaisRequestID": "f73e91c1-6812-5522-14c4-0d762ca66849",
    "id": "http://id.opencalais.com/lI-1l*-GlzU4-tLm773vCg",
    "ontology": "http://och1-lb/owlschema/8.2/onecalais.owl.allmetadata.xml",
    "docId": "http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630",
    "document": "<Document> \n\t<Title>Apple studies self-driving car, auto industry source says</Title>
\n\t<Date>2015-03-09</Date> \n\t<Body> \n(Reuters) - Technology giant Apple (AAPL.O) is looking beyond
mobile devices to learn how to make a self-driving electric car, and is talking to experts at carmakers
and automotive suppliers, a senior auto industry source familiar with the discussions said on Saturday.
\n \nThe Cupertino, California-based maker of phones, computers and, soon, watches is exploring how to
make an entire vehicle, not just designing automotive software or individual components, the auto
industry source said. \n \n"They don't appear to want a lot of help from carmakers,\" said the source,
who declined to be named. \n \nApple is gathering advice on parts and production methods, focusing on
electric and connected-car technologies, while studying the potential for automated driving, the source
said. \n \n\"Fully automated driving is an evolution. Carmakers will slowly build the market for
autonomous cars by first releasing connected and partially automated cars,\" the auto industry source
said. \"Apple is interested in all the potential ways you can evolve the car; that includes autonomous
driving.\" \n \nWhether it will build and release an electric car or a more evolved autonomous vehicle
remains to be seen, the source said. \n \nBut clearly Apple has sharply raised its ambitions in
automotive technology. Car technology has become a prime area of interest for Silicon Valley companies
ranging from Google Inc (GOOGL.O), which has built a prototype self-driving car, to electric car-maker
Tesla Motors Inc (TSLA.O). \n \nAn Apple spokesman in London on Saturday declined to comment on \"rumors
or speculation\". \n \nTrying to build an actual car would mark a dramatic shift for the maker of the
iPhone and iPad. Apple often researches projects which are then discarded, but has so far mainly stuck
to its core expertise in mobile and electronic devices. \n \nThe Wall Street Journal reported on Friday
that Apple had set up a secret lab working on the creation of an Apple-branded electric car, citing
people familiar with the matter. The lab was set up late last year, soon after Apple revealed its
forthcoming smart watch and latest iPhones, the Financial Times said. \n \nThe Journal said that the
Apple project, code-named \"Titan\", employed several hundred people working a few miles from Apple's
headquarters in Cupertino. \n \nApple executives met with contract manufacturers including Magna Steyr
in Austria, a unit of Magna International (MG.TO), the Journal said. A Magna spokeswoman declined to
comment. \n \nTHE PATH TO SELF-DRIVING CARS \n \nAutonomous driving is likely to emerge progressively as
driver assistance systems become more sophisticated. \n \nAlready, carmakers such as Daimler (DAIGn.DE),
BMW (BMWG.DE) and Volkswagen's (VOWG_p.DE) Audi (NSUG.DE) have revealed cars that can travel long
distances without human intervention. \n \nAnalysts at Exane BNP Paribas have said they see a $25
billion market for automated driving technology by 2020, with vehicle intelligence becoming "the key
differentiating factor". But the brokerage does not expect fully automated cars to hit the road until
2025 or 2030, in part due to regulatory hurdles. \n \nShort of building entire cars, there is money to
be made from the software to run a self-driving vehicle, as well as the services associated with
autonomous driving, such as mapping, car-sharing and car recharging services, the auto source said. \n
\n\"It's a software game. It's all about autonomous driving,\" the industry source said. \n \nApple may
be pursuing mainly auto industry expertise rather than full-scale partnerships with established car
companies. \n \nWith its soon-to-be-launched Apple Watch, the company had held limited discussions with
Swiss watchmakers, but no broad-based alliance emerged from the talks. \n \nInstead of partnerships,
Apple pursued a go-it-alone strategy and turned to poaching talent from top watch brands. \n \nTwo
different sources have told Reuters that Apple has tried to recruit auto industry experts in areas such
as robotics. \n \n(Additional reporting by Eric Auchard in Frankfurt; Editing by Noah Barkin/Hugh
Lawson) \n\t</Body> \n</Document>",
    "docTitle": "Apple studies self-driving car, auto industry source says",
    "docDate": "2015-03-09 00:00:00"
},
```

3.4.1.2   Meta

The Meta node presents processing information such as the submission date and time, and the identity of the submitter.

```
"meta": {
        "contentType": "text/xml",
        "processingVer": "AllMetadata",
        "serverVersion": "OneCalais_8.2-RELEASE:360",
        "stagsVer": "OneCalais_8.2-RELEASE-b6-2015-02-14_17:44:23",
        "submissionDate": "2015-03-22 11:36:39.626",
        "submitterCode": "d9d88048-1255-f96b-87c2-22d93db1bd23",
        "signature": "digestalg-
    1|PmYxlNM8avWDTOQjMhMrOmTVPPQ=|gQMTTrBIwMTq0MbDkobw5xmxiO1rVEhBOYs3cvQUVJAE5280L7sOzg==",
        "language": "English"
    }
```

3.4.1.3 Component Versions

The ComponentVersions node specifies the component versions used to process the input file. This information is primarily for the use of the Text Metadata Server (TMS) group at Thomson Reuters, in case of a processing problem.

```
"http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/ComponentVersions": {
        "_typeGroup": "versions",
        "version": ["Deals Index:201503220400:201503220400",
        "OA Index:201503211840:201503211840",
        "NextTags:OneCalais_8.2-RELEASE:108",
        "SpanishIM:OneCalais_8.2-RELEASE:195",
        "config-sca-DataPackage:34:34",
        "SECHeaderMetadataIM:OneCalais_8.2-RELEASE:195",
        "com.clearforest.infoext.dial4j.plugins-basistechconfig:OneCalais_8.2-RELEASE:222",
        "People Index:201503212335:201503212335",
        "Collector:OneCalais_8.2-RELEASE:108",
        "Dial4J:OneCalais_8.2-RELEASE:195",
        "AutocoderRuntimeIM:OneCalais_8.2-RELEASE:195",
        "OA Override:258:258",
        "People Override:247:247",
        "BrokerResearchIM:OneCalais_8.2-RELEASE:195",
        "config-refineries:247:247",
        "config-cse:247:247",
        "OneCalaisIM:OneCalais_8.2-RELEASE:195",
        "config-vessels:247:247",
        "OneCalais:OneCalais_8.2-RELEASE:360",
        "Housekeeper:OneCalais_8.2-RELEASE:108",
        "WatchDog:OneCalais_8.2-RELEASE:108",
        "SocialTags Index:201503080540:201503080540",
        "BlackList:247:247",
        "FrenchIM:OneCalais_8.2-RELEASE:195",
        "config-physicalAssets-ports:247:247",
        "config-drugs:247:247"]
    },
```

3.4.1.4   DefaultLangID

This tag indicates the input text language, specified by the <u>x-calais-language</u> request header, or identified by Open Calais.

Based on the input text language, Open Calais invokes the appropriate metadata extraction module.

In the following example, the indicated input text language is English.

```
"http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/lid/DefaultLangId": {
        "_typeGroup": "language",
        "language": "http://d.opencalais.com/lid/DefaultLangId/English",
        "forenduserdisplay": "false",
        "permid": "505062"
    },
```

Note the *forenduserdisplay* attribute. This is our recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).

3.4.2 Aboutness Tags

Aboutness tags describe the piece of content as a whole.

The Open Calais response may include any of the following metadata tags;

- SocialTag
- DocCat (Topic Tag)
- Industry Tags

3.4.2.1   SocialTag

Social Tags attempt to classify the document as a whole, based on Wikipedia folksonomy.

Examples of social tags, extracted by Open Calais from the article about Apple developing a self-driving car:

```
"http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/SocialTag/1": {
        "_typeGroup": "socialTag",
        "id": "http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/SocialTag/1",
        "socialTag": "http://d.opencalais.com/genericHasher-1/1205cb52-d703-34d2-83b2-a09d4d47575c",
        "forenduserdisplay": "true",
        "name": "Apple Inc.",
        "importance": "1",
        "originalValue": "Apple Inc."
    },
```

SocialTag Attributes

Note the *importance* attribute, which indicates how centric the topic named by the social tag is to the document as a whole. The importance attribute value can be 1 (very centric), 2 (somewhat centric), or 3 (less centric).

The *docID* (dochash) is the unique ID of the containing document.

The Subject (http://d.opencalais.com/dochash-1/268b4e49…) is the tag's unique ID within the containing document. The dochash component of the subject is what associates this tag with this document.

```
"http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/SocialTag/2": {
        "_typeGroup": "socialTag",
        "id": "http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/SocialTag/2",
        "socialTag": "http://d.opencalais.com/genericHasher-1/ccc60460-211d-3b02-b13e-11fba4449fbd",
        "forenduserdisplay": "true",
        "name": "Autonomous car",
        "importance": "1",
        "originalValue": "Autonomous car"
    },
```
SocialTag Attributes

```
"http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/SocialTag/4": {
        "_typeGroup": "socialTag",
        "id": "http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/SocialTag/4",
        "socialTag": "http://d.opencalais.com/genericHasher-1/083d56d1-2fed-3063-b59a-963d4fdaee36",
        "forenduserdisplay": "true",
        "name": "Electric car",
        "importance": "2",
        "originalValue": "Electric car"
    },
```

SocialTag Attributes

3.4.2.2 DocCat (Topic Tag)

Topic tags identify the topic or topics being discussed in the document. The reference list of topics is defined by the Thomson Reuters Coding Schema (TRCS) and/or by the International Press Telecommunications Council (IPTC) news taxonomy and/or by the Self Service Classification project taxonomy. See Supported Topics.

Note that access to TRCS is available upon subscription to Thomson Reuters Intelligent Tagging (TRIT). For more information please contact us at questions@opencalais.com.

Following are examples of topic tags, extracted by Open Calais from the article about Apple developing a self-driving car.

An IPTC taxonomy topic:

```
"http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/cat/1": {
        "_typeGroup": "topics",
        "forenduserdisplay": "false",
        "score": 0.988,
        "name": "Technology_Internet"
    },
```

Topic Tag Attributes

Note the *score* attribute, which indicates the probability, on a scale of 0 to 1, that the topic is indeed discussed in the text, and also how centric the topic is to the text. The higher the value, the higher the probability.

A TRCS taxonomy topic:

```
"http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/cat/2": {
        "_typeGroup": "topics",
        "forenduserdisplay": "false",
        "rcscode": "B:172"
        "name": "Software &amp; IT Services (TRBC)"
        "shortName": "Software &amp; IT Services"
        "score": 0.765,
    },
```

Topic Tag Attributes

3.4.2.3 Industry Tags

Industry tags indicate the industries that are related to the companies mentioned in the input text. The reference list of industries is defined by the Thomson Reuters Business Classification (TRBC) taxonomy.

The following Industry tags were extracted by Open Calais from the story about Apple developing a self-driving car.

```
"http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Industry/5": {
        "_typeGroup": "industry",
        "forenduserdisplay": "false",
        "name": "Automobiles & Multi Utility Vehicles",
        "rcscode": "B:1294",
        "trbccode": "5210101097",
        "permid": "4294951707",
        "relevance": 0.5
    },
```

Industry Tag Attributes

Industry tags include a unique Thomson Reuters ID (the *permid* attribute value). The ID can be used to extract information about the industry from the Thomson Reuters dataset. The ID also supports linkage across documents processed by Open Calais.

Note the *relevance* attribute, which indicates how relevant the industry is to the story. Values range from 0 to 1. The higher the score, the higher the relevance.

```
"http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Industry/7": {
        "_typeGroup": "industry",
        "forenduserdisplay": "false",
        "name": "Electrical (Alternative) Vehicles",
        "rcscode": "B:1296",
        "trbccode": "5210101025",
        "permid": "4294951705",
        "relevance": 0.5
    },
```

Industry Tag Attributes

```
"http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Industry/8": {
        "_typeGroup": "industry",
        "forenduserdisplay": "false",
        "name": "Auto & Truck Manufacturers - NEC",
        "rcscode": "B:1292",
        "trbccode": "5210101029",
        "permid": "4294951709",
        "relevance": 0.5
    },
```

Industry Tag Attributes

3.4.3   Named Entity and Relationship Recognition

Open Calais scans and analyzes the input text, searching for mentions of things like companies, people, deals, and geographical locations, based on a list of predefined metadata types. The resulting tags form the major part of the output response: Instance tags, Entity Markup tags, Relevance tags, Confidence tags, and Disambiguation tags.

In this section we'll highlight some of the metadata tags extracted by Open Calais from the article about Apple developing a self-driving car.

Note that in the JSON output format, all of the tags related to an extracted entity or relation are nested within the entity markup tag.

Example 1: An Extracted Entity of the Type Company

One of the Open Calais predefined entity types is Company. The article mentions several companies. Let's take a look at the metadata tags generated by Open Calais for the company, Apple:

Each group of one or more instances deemed to refer to a unique thing is expressed as an Entity Markup tag. The following is the Entity Markup (em/e/company) tag for the company entity, Apple. (The instances tags, nested within the Entity Markup tag, are also illustrated in this example, further down.)

```
"http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4": {
        "_typeGroup": "entities",
        "_type": "Company",
        "forenduserdisplay": "true",
        "name": "Apple",
        "nationality": "American",
        "confidencelevel": "0.996",
        "_typeReference": "http://s.opencalais.com/1/type/em/e/Company",
```

Em/e/Company Tag Attributes

The hash tag (comphash) is the unique ID assigned by Open Calais to this extracted entity.

The high *confidencelevel* value indicates a high level of confidence that the extracted company, Apple, is indeed a company. Some tag types that include the *confidencelevel* attribute also have an associated Confidence tag.

The associated confidence tag (nested within the Entity Markup tag):

```
        "confidence": {
            "statisticalfeature": "0.997",
            "dblookup": "0.0",
            "resolution": "0.9928677",
            "aggregate": "0.996"
        },
```

Confidence Tag Attributes

The *aggregate* attribute value is the confidence score.

Open Calais succeeded in mapping the extracted company entity, Apple, to the corresponding company and unique ID in the Thomson Reuters dataset, resulting in the following Disambiguation tag (nested within the Entity Markup tag):

```
"resolutions": [{
    "permid": "4295905573",
    "score": 0.9928677,
    "name": "Apple Inc",
    "commonname": "Apple",
    "ticker": "AAPL",
    "primaryric": "AAPL.OQ",
    "id": https://permid.org/1-4295905573,
    "ispublic": "true",
}],
```

Company Disambiguation Tag Attributes

Note: The **id** attribute gives you direct access to high quality, curated Thomson Reuters company data. The attribute value is a direct link to the relevant company page on the Open PermID website (https://permid.org).

The successful mapping to the corresponding Thomson Reuters entity and permid enables the Open Calais entity, in this case the company, Apple, Inc., to be unambiguously identified (and thus linked) across all documents processed by Open Calais.

The following Instance tags were generated when Open Calais found text strings that it identified as mentions of the company, Apple:

"Apple studies self-driving car, auto industry source says"

```
"instances": [{
    "detection": "[<Document> \n\t<Title>]Apple[ studies self-driving car, auto industry source]",
    "prefix": "<Document> \n\t<Title>",
    "exact": "Apple",
    "suffix": " studies self-driving car, auto industry source",
    "offset": 20,
    "length": 5
},
```

Instance Tag Attributes

"Technology giant Apple (APPL.O) is looking beyond mobile devices to learn how to make a self-driving electric car, and is…"

```
{
    "detection": "[\n\t<Body> \n(Reuters) - Technology giant ]Apple (AAPL.0)[ is looking beyond mobile
devices to learn how to]",
    "prefix": "\n\t<Body> \n(Reuters) - Technology giant ",
    "exact": "Apple (AAPL.0)",
    "suffix": " is looking beyond mobile devices to learn how to",
    "offset": 151,
    "length": 14
},
```

Instance Tag Attributes

"An Apple spokesman in London on Saturday declined to comment on 'rumors or speculation.' "

```
{"detection":"[carmaker Tesla Motors Inc(TSLA.0).\n\n An]Apple[spokesman in Londonon Saturday declined to]",
          "prefix":"car-maker Tesla Motors Inc(TSLA.0).\n\n An",
          "exact":"Apple",
          "suffix":"spokesman in Londonon Saturday declined to",
          "offset":1625,
          "length":5
    },
```

Instance Tag Attributes

The RelevanceInfo tag associated with the company entity, Apple, indicates that Apple is highly relevant to the story:

```
    "relevance": 0.8
  },
```

Relevance Tag Attributes

Example 2: An Extracted Relation of the Type CompanyExpansion

The following tags were generated when Open Calais found a mention of a company expansion.

Note that the *company* attribute value is a reference to the extracted company entity, Apple.

```
"http://d.opencalais.com/genericHasher-1/4c3556cd-1c91-363f-b4b2-a9d1b372aa85": {
        "_typeGroup": "relations",
        "_type": "CompanyExpansion",
        "forenduserdisplay": "false",
        "expansiontype": "New Unit",
        "status": "known",
        "_typeReference": "http://s.opencalais.com/1/type/em/r/CompanyExpansion",
        "company": "http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4",

        "instances": [{
            "detection": "[ \nThe Wall Street Journal reported on Friday that ]Apple had set up a secret
        lab[ working on the creation of an Apple-branded]",
            "prefix": " \nThe Wall Street Journal reported on Friday that ",
            "exact": "Apple had set up a secret lab",
            "suffix": " working on the creation of an Apple-branded",
            "offset": 2003,
            "length": 29
        }]
    },
```

em/r/CompanyExpansion Tag Attributes
Instance Tag Attributes

Example 3: An Extracted Relation of the Type CompanyTechnology

The following tags were generated when Open Calais found a mention in the text that it identified as a relationship between a company and a technology.

```
"http://d.opencalais.com/genericHasher-1/55d6ef1d-6140-3150-930c-0d5c2fbd0107": {
        "_typeGroup": "relations",
        "_type": "CompanyTechnology",
        "forenduserdisplay": "false",
        "_typeReference": "http://s.opencalais.com/1/type/em/r/CompanyTechnology",
        "company": "http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4",
        "technology": "http://d.opencalais.com/genericHasher-1/38c58bd0-2536-3f03-bfa0-be1867f6fce8",

        "instances": [{
            "detection": "[\n\t<Date>2015-03-09</Date> \n\t<Body> \n(Reuters) - ]Technology giant Apple (AAPL.0)
        is looking beyond mobile devices[ to learn how to make a self-driving electric]",
            "prefix": "\n\t<Date>2015-03-09</Date> \n\t<Body> \n(Reuters) - ",
            "exact": "Technology giant Apple (AAPL.0) is looking beyond mobile devices",
            "suffix": " to learn how to make a self-driving electric",
            "offset": 134,
            "length": 64
        }]
    },

em/r/CompanyTechnology Tag Attributes
Instance Tag Attributes
```

Note that the company attribute value is the comphash which references the extracted company entity, Apple Inc., and the technology attribute value is the genericHasher which references the extracted technology entity, mobile devices.

Example 4: An Extracted Entity of the Type City

In this example, the Instance, Entity Markup, and Relevance tags were generated when Open Calais found text strings that it identified as references to the same City. The associated Disambiguation (er/Geo/City) tag was generated when Open Calais successfully mapped the extracted city, Cupertino, to the corresponding city in the Thomson Reuters dataset.

Entity markup tag for the extracted City entity, Cupertino:

```
"http://d.opencalais.com/genericHasher-1/752be8ce-c588-3bbe-8526-af3b60708561": {
        "_typeGroup": "entities",
        "_type": "City",
        "forenduserdisplay": "false",
        "name": "Cupertino",
        "_typeReference": "http://s.opencalais.com/1/type/em/e/City",
```

[em/e/City Tag Attributes](#)

Disambiguation tag:

```
        "resolutions": [{
            "name": "Cupertino,California,United States",
            "shortname": "Cupertino",
            "latitude": "37.3231",
            "longitude": "-122.0311",
            "containedbystate": "California",
            "containedbycountry": "United States"
        }],
```

[City Disambiguation Tag Attributes](#)

Instance tags generated for two found mentions of Cupertino in the text.

```
    "instances": [{
        "detection": "[with the discussions said on Saturday. \n \nThe ]Cupertino[, California-based maker
    of phones, computers]",
        "prefix": "with the discussions said on Saturday. \n \nThe ",
        "exact": "Cupertino",
        "suffix": ", California-based maker of phones, computers",
        "offset": 399,
        "length": 9
    },
```

Instance Tag Attributes

```
    {
        "detection": "[ working a few miles from Apple's headquarters in ]Cupertino[. \n \nApple executives
    met with contract]",
        "prefix": " working a few miles from Apple's headquarters in ",
        "exact": "Cupertino",
        "suffix": ". \n \nApple executives met with contract",
        "offset": 2411,
        "length": 9
    }],
```

Instance Tag Attributes

Relevance tag:

```
    "relevance": 0.2
  },
```

Relevance Tag Attributes

The relevance score indicates that the city, Cupertino, is not highly centric to the story.

Example 5: An Extracted Relation of the Type CompanyLocation

The following tags were generated when Open Calais found a text string that it identified as a relationship between a company and a location (i.e. a mention, of the type CompanyLocation):

```
"http://d.opencalais.com/genericHasher-1/f64cef62-b4a7-39f1-b92a-ef8834053ff1": {
        "_typeGroup": "relations",
        "_type": "CompanyLocation",
        "forenduserdisplay": "true",
        "companylocationtype": "N/A",
        "_typeReference": "http://s.opencalais.com/1/type/em/r/CompanyLocation",
        "company": "http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4",
        "city": "http://d.opencalais.com/genericHasher-1/752be8ce-c588-3bbe-8526-af3b60708561",
```

<u>em/r/CompanyLocation Tag Attributes</u>

```
        "instances": [{
            "detection": "[with the discussions said on Saturday. \n \n]The Cupertino, California-based
        maker[ of phones, computers and, soon, watches is]",
            "prefix": "with the discussions said on Saturday. \n \n",
            "exact": "The Cupertino, California-based maker",
            "suffix": " of phones, computers and, soon, watches is",
            "offset": 395,
            "length": 37
        }]
    },
```

<u>Instance Tag Attributes</u>

Note that the *company* attribute value is the comphash that references the extracted company, Apple, Inc., and the *city* attribute value is the genericHasher that references the extracted city, Cupertino.

3.5   N3 Response Format

This section provides an N3 output sample with explanation as a general guideline to parsing and interpreting the Open Calais response.

Note: It is important to note that most attributes are optional; a tag can be extracted with some but not all of its attributes.

The output file is comprised of the following main sections:

- General Document Information
- Aboutness Tags that relate to the document as a whole.
- Named Entity and Relationship Recognition Tags that describe the text strings contained within the document.

3.5.1 General Document Information

[Summary of Extracted Relations and Entities](#)

[DocInfo](#)

[DocInfoMeta](#)

[Component Versions](#)

[DefaultLangID](#)

3.5.1.1 Summary of Extracted Relations and Entities

The output includes a summary of all relations and entities extracted from the text. Extracted entities are displayed alphabetically and grouped by type.

Take a look at the output example, below, and note the following:

- The extracted relations are CompanyExpansion, CompanyLocation, CompanyTechnology, PersonLocation, Quotation.

- The extracted entity types are City, Company, Country, Editor, Facility, IndustryTerm, Journalist, Position, Product, ProvinceOrState, PublishedMedium, Technology.

- The extracted City entities are Cupertino and London; the extracted Company entities are Apple, Audi, BMW, Daimler, etc.; the extracted Country entity is Austria; etc.

```
#Relations: CompanyExpansion, CompanyLocation, CompanyTechnology, PersonLocation, Quotation,
#City: Cupertino, London
#Company: Apple, Audi, BMW, Daimler, Google Inc, Magna International, Reuters, Tesla Motors Inc, Volkswagen,
        Wall Street Journal, the Financial Times
#Country: Austria
#Editor: Hugh Lawson, Noah Barkin
#Facility: Apple's headquarters
#IndustryTerm: Car technology, actual car, automated driving technology, automotive software, automotive
        technology, car, car recharging services, carmakers and automotive suppliers, driver assistance systems,
        electric and connected-car technologies, electric car, mobile and electronic devices, prototype self-
        driving car, self-driving electric car, software game, studies self-driving car
#Journalist: Eric Auchard, Hugh Lawson
#Position: driver, spokesman, spokeswoman
#Product: iPad, iPhone
#ProvinceOrState: California
#PublishedMedium: The Wall Street Journal, the Financial Times
#Technology: Car technology, automated driving technology, electric and connected-car technologies, mobile
        devices
```

3.5.1.2   DocInfo

The DocInfo node presents the original input text (inside the body tag).

Note the dochash, which appears in many of the tags in the output file; it is the unique ID of the containing document.

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630>
        a                    <http://s.opencalais.com/1/type/sys/DocInfo> ;
        c:calaisRequestID    "5bc2bf5e-f634-bfbe-14c4-05297f366849" ;
        c:docDate            "2015-03-09 00:00:00" ;
        c:docTitle           "Apple studies self-driving car, auto industry source says" ;
        c:document           "<Document> \n\t<Title>Apple studies self-driving car, auto industry source
        says</Title> \n\t<Date>2015-03-09</Date> \n\t<Body> \n(Reuters) - Technology giant Apple (AAPL.0) is
        looking beyond mobile devices to learn how to make a self-driving electric car, and is talking to
        experts at carmakers and automotive suppliers, a senior auto industry source familiar with the
        discussions said on Saturday. \n \nThe Cupertino, California-based maker of phones, computers and, soon,
        watches is exploring how to make an entire vehicle, not just designing automotive software or individual
        components, the auto industry source said. \n \n\"They don't appear to want a lot of help from
        carmakers,\" said the source, who declined to be named. \n \nApple is gathering advice on parts and
        production methods, focusing on electric and connected-car technologies, while studying the potential
        for automated driving, the source said. \n \n\"Fully automated driving is an evolution. Carmakers will
        slowly build the market for autonomous cars by first releasing connected and partially automated cars,\"
        the auto industry source said. \"Apple is interested in all the potential ways you can evolve the car;
        that includes autonomous driving.\" \n \nWhether it will build and release an electric car or a more
        evolved autonomous vehicle remains to be seen, the source said. \n \nBut clearly Apple has sharply
        raised its ambitions in automotive technology. Car technology has become a prime area of interest for
        Silicon Valley companies ranging from Google Inc (GOOGL.0), which has built a prototype self-driving
        car, to electric car-maker Tesla Motors Inc (TSLA.0). \n \nAn Apple spokesman in London on Saturday
        declined to comment on \"rumors or speculation\". \n \nTrying to build an actual car would mark a
        dramatic shift for the maker of the iPhone and iPad. Apple often researches projects which are then
        discarded, but has so far mainly stuck to its core expertise in mobile and electronic devices. \n \nThe
        Wall Street Journal reported on Friday that Apple had set up a secret lab working on the creation of an
        Apple-branded electric car, citing people familiar with the matter. The lab was set up late last year,
        soon after Apple revealed its forthcoming smart watch and latest iPhones, the Financial Times said. \n
        \nThe Journal said that the Apple project, code-named \"Titan\", employed several hundred people working
        a few miles from Apple's headquarters in Cupertino. \n \nApple executives met with contract
        manufacturers including Magna Steyr in Austria, a unit of Magna International (MG.TO), the Journal said.
        A Magna spokeswoman declined to comment. \n \nTHE PATH TO SELF-DRIVING CARS \n \nAutonomous driving is
        likely to emerge progressively as driver assistance systems become more sophisticated. \n \nAlready,
        carmakers such as Daimler (DAIGn.DE), BMW (BMWG.DE) and Volkswagen's (VOWG_p.DE) Audi (NSUG.DE) have
        revealed cars that can travel long distances without human intervention. \n \nAnalysts at Exane BNP
        Paribas have said they see a $25 billion market for automated driving technology by 2020, with vehicle
        intelligence becoming "the key differentiating factor". But the brokerage does not expect fully
        automated cars to hit the road until 2025 or 2030, in part due to regulatory hurdles. \n \nShort of
        building entire cars, there is money to be made from the software to run a self-driving vehicle, as well
        as the services associated with autonomous driving, such as mapping, car-sharing and car recharging
        services, the auto source said. \n \n\"It's a software game. It's all about autonomous driving,\" the
        industry source said. \n \nApple may be pursuing mainly auto industry expertise rather than full-scale
        partnerships with established car companies. \n \nWith its soon-to-be-launched Apple Watch, the company
        had held limited discussions with Swiss watchmakers, but no broad-based alliance emerged from the talks.
        \n \nInstead of partnerships, Apple pursued a go-it-alone strategy and turned to poaching talent from
        top watch brands. \n \nTwo different sources have told Reuters that Apple has tried to recruit auto
        industry experts in areas such as robotics. \n \n(Additional reporting by Eric Auchard in Frankfurt;
        Editing by Noah Barkin/Hugh Lawson) \n\t</Body> \n</Document>" ;
        c:id                 "http://id.opencalais.com/lI-1l*-GlzU4-tLm773vCg" ;
        c:ontology           "http://och1-lb/owlschema/8.2/onecalais.owl.allmetadata.xml" .
```

3.5.1.3 DocInfoMeta

The DocInfoMeta node presents processing information such as the submission date and time, and the identity of the submitter.

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/meta>
        a                  <http://s.opencalais.com/1/type/sys/DocInfoMeta> ;
        c:contentType      "text/xml" ;
        c:docId            <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:language         "English" ;
        c:processingVer    "AllMetadata" ;
        c:serverVersion    "OneCalais_8.2-RELEASE:360" ;
        c:signature        "digestalg-
        1|ybwQVZp7FTG8o10IvznzbffxKsg=|i/uDr/e/Kg73W4v8PyPROSQGK7OuxQvrAObJloQkcBVCzUUYVxhgWQ==" ;
        c:stagsVer         "OneCalais_8.2-RELEASE-b6-2015-02-14_17:44:23" ;
        c:submissionDate   "2015-03-22 09:11:36.947" ;
        c:submitterCode    "d9d88048-1255-f96b-87c2-22d93db1bd23" .
```

3.5.1.4   Component Versions

The ComponentVersions node specifies the component versions used to process the input file. This information is primarily for the use of the Text Metadata Server (TMS) group at Thomson Reuters, in case of a processing problem.

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/ComponentVersions>
        a              <http://s.opencalais.com/1/type/sys/ComponentVersions> ;
        c:docId        <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:version   "People Override:247:247" , "OA Override:258:258" , "OA Index:201503211840:201503211840" ,
        "Dial4J:OneCalais_8.2-RELEASE:195" , "NextTags:OneCalais_8.2-RELEASE:108" , "People
        Index:201503212335:201503212335" , "config-vessels:247:247" , "FrenchIM:OneCalais_8.2-RELEASE:195" ,
        "Collector:OneCalais_8.2-RELEASE:108" , "config-drugs:247:247" , "WatchDog:OneCalais_8.2-RELEASE:108" ,
        "SocialTags Index:201503080540:201503080540" , "Housekeeper:OneCalais_8.2-RELEASE:108" , "config-
        refineries:247:247" , "AutocoderRuntimeIM:OneCalais_8.2-RELEASE:195" , "OneCalaisIM:OneCalais_8.2-
        RELEASE:195" , "OneCalais:OneCalais_8.2-RELEASE:360" , "SECHeaderMetadataIM:OneCalais_8.2-RELEASE:195" ,
        "BlackList:247:247" , "BrokerResearchIM:OneCalais_8.2-RELEASE:195" , "SpanishIM:OneCalais_8.2-
        RELEASE:195" , "config-cse:247:247" , "Deals Index:201503220400:201503220400" , "config-sca-
        DataPackage:34:34" , "com.clearforest.infoext.dial4j.plugins-basistechconfig:OneCalais_8.2-
        RELEASE:222" , "config-physicalAssets-ports:247:247" .
```

3.5.1.5 DefaultLangID

This tag indicates the input text language, specified by the <u>x-calais-language</u> request header, or identified by Open Calais.

Based on the input text language, Open Calais invokes the appropriate metadata extraction module.

In the following example, the indicated input text language is English.

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/lid/DefaultLangId>
        a                       <http://s.opencalais.com/1/type/lid/DefaultLangId> ;
        c:docId                 <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:forenduserdisplay     "false" ;
        c:lang                  <http://d.opencalais.com/lid/DefaultLangId/English> ;
        c:permid                "505062" .
```

Note the *forenduserdisplay* attribute. This is our recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).

3.5.2   Aboutness Tags

Aboutness tags describe the piece of content as a whole.

The Open Calais response may include any of the following metadata tags:

- SocialTag
- DocCat (Topic Tags)
- Industry Tags

3.5.2.1 SocialTag

Social tags attempt to classify the document as a whole, based on Wikipedia folksonomy.

Examples of social tags, extracted by Open Calais from the article about Apple developing a self-driving car:

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/SocialTag/1>
        a                      <http://s.opencalais.com/1/type/tag/SocialTag> ;
        c:docId                <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:forenduserdisplay    "true" ;
        c:importance           "1" ;
        c:name                 "Apple Inc." ;
        c:originalValue        "Apple Inc." ;
        c:socialtag            <http://d.opencalais.com/genericHasher-1/1205cb52-d703-34d2-83b2-a09d4d47575c>
```

SocialTag Attributes

Note the *importance* attribute, which indicates how centric the topic named by the social tag is to the document as a whole. The importance attribute value can be 1 (very centric), 2 (somewhat centric), or 3 (less centric).

The *docID* (dochash) is the unique ID of the containing document.

The Subject (http://d.opencalais.com/dochash. . .) is the tag's unique ID within the containing document. The dochash component of the Subject is what associates this tag with this document.

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/SocialTag/2>
        a                      <http://s.opencalais.com/1/type/tag/SocialTag> ;
        c:docId                <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:forenduserdisplay    "true" ;
        c:importance           "1" ;
        c:name                 "Autonomous car" ;
        c:originalValue        "Autonomous car" ;
        c:socialtag            <http://d.opencalais.com/genericHasher-1/ccc60460-211d-3b02-b13e-11fba4449fbd> .
```

SocialTag Attributes

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/SocialTag/4>
        a                      <http://s.opencalais.com/1/type/tag/SocialTag> ;
        c:docId                <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:forenduserdisplay    "true" ;
        c:importance           "2" ;
        c:name                 "Electric car" ;
        c:originalValue        "Electric car" ;
        c:socialtag            <http://d.opencalais.com/genericHasher-1/083d56d1-2fed-3063-b59a-963d4fdaee36> .
```

SocialTag Attributes

3.5.2.2　DocCat (Topic Tag)

Topic tags identify the topic or topics being discussed in the document. The reference list of topics is defined by the Thomson Reuters Coding Schema (TRCS) and/or by the International Press Telecommunications Council (IPTC) news taxonomy and/or by the Calais Self Service Classification project taxonomy. See Supported Topics.

Note that access to TRCS is available upon subscription to Thomson Reuters Intelligent Tagging (TRIT). For more information please contact us at questions@opencalais.com.

Following are examples of topic tags, extracted by Open Calais from the article about Apple developing a self-driving car.

An IPTC taxonomy topic:

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/cat/1>
        a                       <http://s.opencalais.com/1/type/cat/DocCat> ;
        c:docId                 <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:forenduserdisplay     "false" ;
        c:name                  "Technology_Internet" ;
        c:score                 "0.988" .
```

Topic Tag (DocCat tag) attributes

Note the *score* attribute, which indicates the probability, on a scale of 0 to 1, that the topic is indeed discussed in the text, and also how centric the topic is to the text; the higher the value, the higher the probability.

A TRCS taxonomy topic:

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/cat/2>
        a                       <http://s.opencalais.com/1/type/cat/DocCat> ;
        c:docId                 <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:forenduserdisplay     "false" ;
        c:rcscode               "B:172" ;
        c:score                 "0.765" ;
        c:name                  "Software &amp; IT Services (TRBC)" ;
        c:shortName             "Software &amp; IT Services" .
```

Topic Tag (DocCat tag) attributes

3.5.2.3   Industry Tags

Industry tags indicate the industries that are related to the companies mentioned in the input text. The reference list of industries is defined by the Thomson Reuters Business Classification (TRBC) taxonomy.

The following Industry tags were extracted by Open Calais from the story about Apple developing a self-driving car.

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Industry/5>
        a                      <http://s.opencalais.com/1/type/tag/Industry> ;
        c:docId                <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:forenduserdisplay    "false" ;
        c:name                 "Automobiles & Multi Utility Vehicles" ;
        c:permid               "4294951707" ;
        c:trbccode             "5210101097" ;
        c:rcscode              "B:1294" ;
        c:relevance            "0.500" .
```

Industry Tag Attributes

Industry tags include a unique Thomson Reuters ID (the *permid* attribute value). The ID can be used to extract information about the industry from the Thomson Reuters dataset. The ID also supports linkage across documents processed by Open Calais.

Note the *relevance* attribute, which indicates how relevant the industry is to the story. Values range from 0 to 1; the higher the score, the higher the relevance.


```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Industry/7>
        a                      <http://s.opencalais.com/1/type/tag/Industry> ;
        c:docId                <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:forenduserdisplay    "false" ;
        c:name                 "Electrical (Alternative) Vehicles" ;
        c:permid               "4294951705" ;
        c:trbccode             "5210101025" ;
        c:rcscode              "B:1296" ;
        c:relevance            "0.500" .
```

Industry Tag Attributes


```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Industry/8>
        a                      <http://s.opencalais.com/1/type/tag/Industry> ;
        c:docId                <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:forenduserdisplay    "false" ;
        c:name                 "Auto & Truck Manufacturers - NEC" ;
        c:permid               "4294951709" ;
        c:trbccode             "5210101029" ;
        c:rcscode              "B:1292" ;
        c:relevance            "0.500" .
```

Industry Tag Attributes

3.5.3　Named Entity and Relationship Recognition

Open Calais scans and analyzes the input text, searching for mentions of things like companies, people, deals, and geographical locations, based on a <u>list of predefined metadata types</u>. The resulting tags form the major part of the output response: <u>Instance tags</u>, <u>Entity Markup tags</u>, <u>Relevance tags</u>, <u>Confidence tags</u>, and <u>Disambiguation tags</u>.

In this section we'll highlight some of the metadata tags extracted by Open Calais from the article about Apple developing a self-driving car.

One of the Open Calais predefined metadata types is Company. The article mentions several companies. Let's take a look at the InstanceInfo tags generated by Open Calais for a few of the found mentions of the company, Apple:

"Apple studies self-driving car, auto industry source says"

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Instance/59>
        a             <http://s.opencalais.com/1/type/sys/InstanceInfo> ;
        c:detection   "[<Document> \n\t<Title>]Apple[ studies self-driving car, auto industry source]" ;
        c:docId       <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:exact       "Apple" ;
        c:length      "5" ;
        c:offset      "20" ;
        c:prefix      "<Document> \n\t<Title>" ;
        c:subject     <http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4> ;
        c:suffix      " studies self-driving car, auto industry source" .
```

<u>InstanceInfo Tag Attributes</u>

"Technology giant Apple (APPL.O) is looking beyond mobile devices to learn how to make a self-driving electric car, and is…"

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Instance/60>
        a             <http://s.opencalais.com/1/type/sys/InstanceInfo> ;
        c:detection   "[\n\t<Body> \n(Reuters) - Technology giant ]Apple (AAPL.O)[ is looking beyond mobile
        devices to learn how to]" ;
        c:docId       <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:exact       "Apple (AAPL.O)" ;
        c:length      "14" ;
        c:offset      "151" ;
        c:prefix      "\n\t<Body> \n(Reuters) - Technology giant " ;
        c:subject     <http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4> ;
        c:suffix      " is looking beyond mobile devices to learn how to" .
```

<u>InstanceInfo Tag Attributes</u>

"An Apple spokesman in London on Saturday declined to comment on 'rumors or speculation.' "

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Instance/63>
        a               <http://s.opencalais.com/1/type/sys/InstanceInfo> ;
        c:detection    "[car-maker Tesla Motors Inc (TSLA.0). \n \nAn ]Apple[ spokesman in London on Saturday
        declined to]" ;
        c:docId        <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:exact        "Apple" ;
        c:length       "5" ;
        c:offset       "1625" ;
        c:prefix       "car-maker Tesla Motors Inc (TSLA.0). \n \nAn " ;
        c:subject      <http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4> ;
        c:suffix       " spokesman in London on Saturday declined to" .
```

InstanceInfo Tag Attributes

"The Journal said that the project, code-named 'Titan,' employed several hundred people working a few miles from Apple's headquarters in Cupertino."

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Instance/68>
        a               <http://s.opencalais.com/1/type/sys/InstanceInfo> ;
        c:detection    "[several hundred people working a few miles from ]Apple['s headquarters in Cupertino. \n
        \nApple executives]" ;
        c:docId        <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:exact        "Apple" ;
        c:length       "5" ;
        c:offset       "2387" ;
        c:prefix       "several hundred people working a few miles from " ;
        c:subject      <http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4> ;
        c:suffix       "'s headquarters in Cupertino. \n \nApple executives" .
```

InstanceInfo Tag Attributes

Each group of one or more instances deemed to refer to a unique thing is expressed as an Entity Markup tag. The following example is the Entity Markup (em/e/company) tag for the company entity, Apple.

The hash tag (comphash) is the unique ID assigned by Open Calais to this extracted entity. This hash tag also appears in the associated instance tags.

```
<http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4>
        a                    <http://s.opencalais.com/1/type/em/e/Company> ;
        c:confidencelevel    "0.996" ;
        c:forenduserdisplay  "true" ;
        c:name               "Apple" ;
        c:nationality        "American" .
```

em/e/Company Tag Attributes

The high *confidencelevel* value indicates a high level of confidence that the extracted company, Apple, is indeed a company. Some tag types that include the *confidencelevel* attribute also have an associated Confidence tag.

The associated Confidence tag:

```
<http://d.opencalais.com/conf/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4>
        a                       <http://s.opencalais.com/1/type/tag/Confidence> ;
        c:aggregate             "0.996" ;
        c:dblookup              "0.0" ;
        c:docId                 <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:resolution            "0.9928677" ;
        c:statisticalfeature    "0.997" ;
        c:subject               <http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4> .
```

<u>Confidence Tag Attributes</u>

Again, note that the entity markup tag and all of its associated tags display the same comphash.

The *aggregate* attribute value is the confidence score.

The RelevanceInfo tag associated with the company entity, Apple, indicates that Apple is highly relevant to the story:

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Relevance/47>
        a                   <http://s.opencalais.com/1/type/sys/RelevanceInfo> ;
        c:docId             <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:relevance         "0.8" ;
        c:relevancecont     "0.72" ;
        c:subject           <http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4> .
```

<u>RelevanceInfo Tag Attributes</u>

Compare the RelevanceInfo tag for the extracted company, Apple, with the Relevance tag associated with the extracted company, Daimler, which was mentioned in the document, along with several other companies, as having "…revealed cars that can travel long distances without human intervention….:"

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Relevance/36>
        a                   <http://s.opencalais.com/1/type/sys/RelevanceInfo> ;
        c:docId             <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:relevance         "0.5" ;
        c:relevancecont     "0.568" ;
        c:subject           <http://d.opencalais.com/comphash-1/ebd24b16-1574-33c0-9b69-30c054337d44> .
```

<u>RelevanceInfo Tag Attributes</u>

Note that the comphash assigned to Daimler is different from the comphash assigned to Apple.

Open Calais succeeded in mapping the extracted company entity, Apple, to the corresponding company and unique ID in the Thomson Reuters dataset, resulting in the following Disambiguation (er/company) tag:

```
<http://d.opencalais.com/er/company/ralg-oa/4295905573>
        a               <http://s.opencalais.com/1/type/er/Company> ;
        c:commonname    "Apple" ;
        c:docId         <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:legacyid      <http://d.opencalais.com/er/company/ralg-tr1r/23d07771-c50b-315b-8050-3cdaf47ac0d0> ;
        c:name          "Apple Inc" ;
        c:openpermid    <https://permid.org/1-4295905573> ;
        c:permid        "4295905573" ;
        c:primaryric    "AAPL.OQ" ;
        c:score         "0.9928677" ;
        c:subject       <http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4> ;
        c:ticker        "AAPL" ;
        c:ispublic      "true" .
```

er/company Tag Attributes

The comphash assigned by Open Calais identifies and links all the associated tags within the output document (local disambiguation).

The successful mapping to the corresponding Thomson Reuters entity and permid enables the Open Calais entity, in this case the company, Apple, Inc., to be unambiguously identified (and thus linked) across all documents processed by Open Calais.

Note: The **openpermid** attribute gives you direct access to high quality, curated Thomson Reuters company data. The attribute value is a direct link to the relevant company page on the Open PermID website (https://permid.org).

The following instance tag was generated when Open Calais found a mention of a company expansion:

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Instance/18>
        a               <http://s.opencalais.com/1/type/sys/InstanceInfo> ;
        c:detection     "[ \nThe Wall Street Journal reported on Friday that ]Apple had set up a secret
        lab[ working on the creation of an Apple-branded]" ;
        c:docId         <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:exact         "Apple had set up a secret lab" ;
        c:length        "29" ;
        c:offset        "2003" ;
        c:prefix        " \nThe Wall Street Journal reported on Friday that " ;
        c:subject       <http://d.opencalais.com/genericHasher-1/4c3556cd-1c91-363f-b4b2-a9d1b372aa85> ;
        c:suffix        " working on the creation of an Apple-branded" .
```

InstanceInfo Tag Attributes

The resulting Entity Markup (em/r/CompanyExpansion) tag (the extracted relation):

```
<http://d.opencalais.com/genericHasher-1/4c3556cd-1c91-363f-b4b2-a9d1b372aa85>
        a                    <http://s.opencalais.com/1/type/em/r/CompanyExpansion> ;
        c:company            <http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4> ;
        c:expansiontype      "New Unit" ;
        c:forenduserdisplay  "false" ;
        c:status             "known" .

em/r/CompanyExpansion Tag Attributes
```

Note that the genericHasher links this Entity Markup tag to the associated InstanceInfo tag.

Also note that the *company* attribute value is a reference to the extracted company entity, Apple.

The following instance tag was generated when Open Calais found a text string that it identified as an indication of a relationship between a company and a technology (i.e. a mention of the type CompanyTechnology):

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Instance/26>
        a               <http://s.opencalais.com/1/type/sys/InstanceInfo> ;
        c:detection  "[\n\t<Date>2015-03-09</Date> \n\t<Body> \n(Reuters) - ]Technology giant Apple (AAPL.0) is
        looking beyond mobile devices[ to learn how to make a self-driving electric]" ;
        c:docId      <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:exact      "Technology giant Apple (AAPL.0) is looking beyond mobile devices" ;
        c:length     "64" ;
        c:offset     "134" ;
        c:prefix     "\n\t<Date>2015-03-09</Date> \n\t<Body> \n(Reuters) - " ;
        c:subject    <http://d.opencalais.com/genericHasher-1/55d6ef1d-6140-3150-930c-0d5c2fbd0107> ;
        c:suffix     " to learn how to make a self-driving electric" .

InstanceInfo Tag Attributes
```

The resulting Entity Markup (em/r/CompanyTechnology) tag:

```
<http://d.opencalais.com/genericHasher-1/55d6ef1d-6140-3150-930c-0d5c2fbd0107>
        a                    <http://s.opencalais.com/1/type/em/r/CompanyTechnology> ;
        c:company            <http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4> ;
        c:forenduserdisplay  "false" ;
        c:technology         <http://d.opencalais.com/genericHasher-1/38c58bd0-2536-3f03-bfa0-be1867f6fce8> .

em/r/CompanyTechnology Tag Attributes
```

Again, note the genericHasher that links this Entity Markup tag to the associated InstanceInfo tag, and note that the company attribute value is the comphash which identifies the company, Apple Inc. The *technology* attribute value references the relevant em/e/technology entity, not illustrated here.

The following instance tag was generated when Open Calais found a text string that it identified as a reference to a City:

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Instance/35>
        a              <http://s.opencalais.com/1/type/sys/InstanceInfo> ;
        c:detection   "[with the discussions said on Saturday. \n \nThe ]Cupertino[, California-based maker of
        phones, computers]" ;
        c:docId        <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:exact        "Cupertino" ;
        c:length       "9" ;
        c:offset       "399" ;
        c:prefix       "with the discussions said on Saturday. \n \nThe " ;
        c:subject      <http://d.opencalais.com/genericHasher-1/752be8ce-c588-3bbe-8526-af3b60708561> ;
        c:suffix       ", California-based maker of phones, computers" .
```

InstanceInfo Tag Attributes

The resulting Entity Markup (em/e/city) tag:

```
<http://d.opencalais.com/genericHasher-1/752be8ce-c588-3bbe-8526-af3b60708561>
        a                    <http://s.opencalais.com/1/type/em/e/City> ;
        c:forenduserdisplay  "false" ;
        c:name               "Cupertino" .
```

em/e/City Tag Attributes

The associated Disambiguation (er/Geo/City) tag was generated when Open Calais succeeded in mapping the extracted city entity, Cupertino, to the corresponding city in the Thomson Reuters dataset:

```
<http://d.opencalais.com/er/geo/city/ralg-geo1/4edd5509-ee4d-deba-738c-fc3900017f64>
        a                    <http://s.opencalais.com/1/type/er/Geo/City> ;
        c:containedbycountry "United States" ;
        c:containedbystate   "California" ;
        c:docId              <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:latitude           "37.3231" ;
        c:longitude          "-122.0311" ;
        c:name               "Cupertino,California,United States" ;
        c:shortname          "Cupertino" ;
        c:subject            <http://d.opencalais.com/genericHasher-1/752be8ce-c588-3bbe-8526-af3b60708561> .
```

er/Geo/City Tag Attributes

The following InstanceInfo tag was generated when Open Calais found a text string that it identified as an indication of a relationship between a company and a location (i.e. a mention, of the type CompanyLocation):

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Instance/32>
        a               <http://s.opencalais.com/1/type/sys/InstanceInfo> ;
        c:detection   "[with the discussions said on Saturday. \n \n]The Cupertino, California-based maker[ of
        phones, computers and, soon, watches is]" ;
        c:docId         <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:exact         "The Cupertino, California-based maker" ;
        c:length        "37" ;
        c:offset        "395" ;
        c:prefix        "with the discussions said on Saturday. \n \n" ;
        c:subject       <http://d.opencalais.com/genericHasher-1/f64cef62-b4a7-39f1-b92a-ef8834053ff1> ;
        c:suffix        " of phones, computers and, soon, watches is" .
```

[InstanceInfo Tag Attributes](#)

The resulting Entity Markup (em/r/CompanyLocation) tag:

```
<http://d.opencalais.com/genericHasher-1/f64cef62-b4a7-39f1-b92a-ef8834053ff1>
        a                       <http://s.opencalais.com/1/type/em/r/CompanyLocation> ;
        c:city                  <http://d.opencalais.com/genericHasher-1/752be8ce-c588-3bbe-8526-af3b60708561> ;
        c:company               <http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4> ;
        c:companylocationtype   "N/A" ;
        c:forenduserdisplay     "true" ;
        c:provinceorstate       <http://d.opencalais.com/genericHasher-1/9679b237-33e8-3478-ba13-d9af3c4b943e> .
```

[em/r/CompanyLocation Tag Attributes](#)

The genericHasher in the Subject (http://d.opencalais.com/genericHasher. . .) links this tag to the associated InstanceInfo tag; the *company* attribute value is the comphash that identifies Apple, Inc.; the *city* attribute value is the genericHasher that identifies the city, Cupertino.

3.6 Error Messages

Open Calais returns messages when it cannot process or complete a transaction due to format or load issues. Such requests are returned in the body of the HTTP response with an HTTP error code.

There are two types of error messages:

- Client Errors (4XX)
- Server Errors (5XX)

3.6.1   Client Errors

If the submitted request contains an error, a client error message is generated; check the message and adapt the request according to the error message.

HTTP Response Code	Error Reason	Error Message	Remarks
400	Bad content-type header value: <VALUE>		
400	Invalid content	Null or empty content submitted.	
400	Unsupported-Language	You've submitted a document in <LANGUAGE>, which is not currently supported.	
400	Unrecognized-Language	Unrecognized language.	Open Calais may not properly identify the language if the input text is too short. In this case, define a valid x-calais-language input header.
401	Invalid API key	oauth.v2.InvalidApiKey	Indicates an invalid license key. Please verify that you have entered the correct key, or contact support.
401	Invalid API key for given resource	oauth.v2.InvalidApiKeyForGivenResource	The license key does not match the POST URL. Please check that the POST address and license key are correct.
406	Unsupported-Output-Format-Requested	Unsupported output format: <VALUE>	
413	Request Entity Too Large	Your Calais request has exceeded the max allowed document size.	
415	Unsupported Media Type	javax.ws.rs.WebApplicationException.	The input format sent is not supported
429	Too many requests	You exceeded the concurrent request limit for your license key. Please try again later or contact support to upgrade your license.	This error is recoverable. Try resubmitting the document (maximum three retries) with a sleep of 750 milliseconds between resubmissions.)
429	Too many requests	You exceeded the allowed quota of <VALUE> requests per day. Please try again at: <DATE_TIME>	This error is recoverable, but is an indication that you have reached your daily quota. Wait to resubmit the document at least until the next day at 00:00:00 GMT time. To increase the quota (the daily document upload limit), please contact us at questions@opencalais.com

3.6.2 Server Errors

Server errors are generated when there is a problem at the server; resubmitting the request might solve the problem.

HTTP Response Code	Error Reason	Error Message	Remarks
500	Request-Timeout	Timeout reached while processing the document.	This error may be recoverable, if the timeout was due to heavy utilization of the system. Try resubmitting the document (maximum three retries) with a sleep of 750 milliseconds between resubmissions.) If the resubmission does not work, it may be that the input document is too complex (contains too many entities and relations) to be processed within the defined time limit. In this case, you can try splitting the document into smaller parts for processing.
500	Request-Terminated	Calais server terminated the job.	
500	Internal-Error	<The actual exception from the server.>	
503	Server-Too-Busy	Calais server is busy. Please try again later.	This error is recoverable. Try resubmitting the document (maximum three retries) with a sleep of 750 milliseconds between resubmissions.)
503	Service Unavailable	No server is available to handle this request.	

Chapter 4   Semantic Metadata Tags

Open Calais is able to extract a rich set of metadata tags from your input text. The following sections provide detailed information about each of the tags, including a list of tag attributes, and examples.

Instance Tag

Entity Markup Tags (Entities, Relations)

RelevanceInfo Tag

Confidence Tag

Disambiguation Tags

SocialTag

Topic Tag (DocCat)

Industry Tag

French Language Support

Spanish Language Support

4.1 InstanceInfo Tag

For a conceptual explanation of this tag, see Instance Tags.

InstanceInfo	
Definition	Describes a mention of an Open Calais type found in the text.
Attributes	**detection**: The text string in which the mention was identified.
	docid: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	exact: The mention.
	length: The length (in characters) of the mention.
	offset: Offset of the mention (in characters) from the beginning of the document.
	prefix: The portion of the text string that precedes the mention.
	subject: A hash tag generated by Open Calais. This ID points to the associated extracted entity or relation.
	suffix: The portion of the text string that follows the mention.
	The tag also includes a comment that indicates the type of entity or relation extracted from this instance, and its main attribute values.
	Note: In the JSON output format, because related tags are nested, the docid and subject attributes are not displayed in the Instance tag.

Example 1: A mention of a Company (Company type)

Input Text	Technology giant Apple (APPL.O) is looking beyond mobile devices to learn how to make a self-driving electric car.

Output Instance Tag	*RDF...*
	
	 ```<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6/Instance/60">```
	``` <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/InstanceInfo"/>```
	``` <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>```
	``` <c:subject rdf:resource="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4"/>```
	``` <!--Company: Apple; -->```
	``` <c:detection>[<Body> (Reuters) - Technology giant]Apple (AAPL.0)[is looking beyond mobile devices to learn how to]</c:detection>```
	``` <c:prefix> <Body> (Reuters) - Technology giant </c:prefix>```
	``` <c:exact>Apple (AAPL.0)</c:exact>```
	``` <c:suffix> is looking beyond mobile devices to learn how to</c:suffix>```
	``` <c:offset>151</c:offset>```
	``` <c:length>14</c:length>```
	``` </rdf:Description>```
	
	 JSON...
	
	 ``` {```
	``` "detection": "[\n\t<Body> \n(Reuters) - Technology giant]Apple (AAPL.0)[is looking beyond mobile devices to learn how to]",```
	``` "prefix": "\n\t<Body> \n(Reuters) - Technology giant ",```
	``` "exact": "Apple (AAPL.0)",```
	``` "suffix": " is looking beyond mobile devices to learn how to",```
	
	``` "offset": 151,```
	``` "length": 14```
	``` },```
	
	 N3…
	
	 ```<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Instance/60>```
	``` a <http://s.opencalais.com/1/type/sys/InstanceInfo> ;```
	``` c:detection "[\n\t<Body> \n(Reuters) - Technology giant]Apple (AAPL.0)[is looking beyond mobile devices to learn how to]" ;```
	``` c:docId <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;```
	``` c:exact "Apple (AAPL.0)" ;```
	``` c:length "14" ;```
	``` c:offset "151" ;```
	``` c:prefix "\n\t<Body> \n(Reuters) - Technology giant " ;```
	``` c:subject <http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4> ;```
	``` c:suffix " is looking beyond mobile devices to learn how to" .```

Example 2: A mention of a company location (CompanyLocation type)

Input Text	Technology giant Apple (AAPL.O) is looking beyond mobile devices to learn how to make a self-driving electric car, and is talking to experts at carmakers and automotive suppliers, a senior auto industry source familiar with the discussions said on Saturday. The Cupertino, California-based maker of phones, computers and, soon, watches is exploring how to make an entire vehicle, not just designing automotive software or individual components, the auto industry source said.
Output Instance Tag	*RDF...*

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
        bf3520cddfa6/Instance/33">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/InstanceInfo"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
        bf3520cddfa6"/>
        <c:subject rdf:resource="http://d.opencalais.com/genericHasher-1/f64cef62-b4a7-39f1-
        b92a-ef8834053ff1"/>
        <!--CompanyLocation: company: Apple; companylocationtype: N/A; city: Cupertino;
        provinceorstate: California; -->
        <c:detection>[with the discussions said on Saturday.]The Cupertino, California-based
        maker[ of phones, computers and, soon, watches is]</c:detection>
        <c:prefix>with the discussions said on Saturday. </c:prefix>
        <c:exact>The Cupertino, California-based maker</c:exact>
        <c:suffix> of phones, computers and, soon, watches is</c:suffix>
        <c:offset>395</c:offset>
        <c:length>37</c:length>
    </rdf:Description>
```

JSON...

```
        "instances": [{
            "detection": "[with the discussions said on Saturday. \n \n]The Cupertino,
        California-based maker[ of phones, computers and, soon, watches is]",
            "prefix": "with the discussions said on Saturday. \n \n",
            "exact": "The Cupertino, California-based maker",
            "suffix": " of phones, computers and, soon, watches is",
            "offset": 395,
            "length": 37
        }]
    },
```

N3...

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Instance/32>
        a              <http://s.opencalais.com/1/type/sys/InstanceInfo> ;
        c:detection  "[with the discussions said on Saturday. \n \n]The Cupertino, California-
        based maker[ of phones, computers and, soon, watches is]" ;
        c:docId       <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
        c:exact       "The Cupertino, California-based maker" ;
        c:length      "37" ;
        c:offset      "395" ;
        c:prefix      "with the discussions said on Saturday. \n \n" ;
        c:subject     <http://d.opencalais.com/genericHasher-1/f64cef62-b4a7-39f1-b92a-
        ef8834053ff1> ;
        c:suffix      " of phones, computers and, soon, watches is" .
```

4.2   Entity Markup Tags

Open Calais metadata types are classified as Entities or Relations. Entities are straightforward things like companies, people, cities, telephone numbers; Relations are more complex mentions that indicate relationships between things. For example, deals, IPOs, analyst recommendations, company reorganizations.

For a conceptual explanation, see Entity Markup Tags.

Click a link for a description of the entity type and its attributes.

Note: It is important to note that most attributes are optional; a tag can be extracted with some but not all of its attributes.

Open Calais Entities:

Anniversary, City, Company, Continent, Country, Currency, Editor, EmailAddress, EntertainmentAwardEvent, Facility, FaxNumber, Holiday, IndustryTerm, Journalist, MarketIndex, MedicalCondition, MedicalTreatment, Movie, MusicAlbum, MusicGroup, NaturalFeature, OperatingSystem, Organization, Person, PharmaceuticalDrug, PhoneNumber, PoliticalEvent, Position, Product, ProgrammingLanguage, ProvinceOrState, PublishedMedium, RadioProgram, RadioStation, Region, SportsEvent, SportsGame, SportsLeague, Technology, TVShow, TVStation, URL

Open Calais Relations:

Acquisition, Alliance, AnalystEarningsEstimate, AnalystRecommendation, ArmedAttack, ArmsPurchaseSale, Arrest, Bankruptcy, BonusSharesIssuance, BusinessRelation, Buybacks, CandidatePosition, CompanyAccountingChange, CompanyAffiliates, CompanyCompetitor, CompanyCustomer, CompanyEarningsAnnouncement, CompanyEarningsGuidance, CompanyEmployeesNumber, CompanyExpansion, CompanyForceMajeure, CompanyFounded, CompanyInvestigation, CompanyInvestment, CompanyLaborIssues, CompanyLayoffs, CompanyLegalIssues, CompanyListingChange, CompanyLocation, CompanyMeeting, CompanyNameChange, CompanyProduct, CompanyReorganization, CompanyRestatement, CompanyTechnology, CompanyTicker, CompanyUsingProduct, ConferenceCall, ContactDetails, Conviction, CreditRating, Deal, DebtFinancing, DelayedFiling, DiplomaticRelations, Dividend, EmploymentChange, EmploymentRelation, EnvironmentalIssue, EquityFinancing, Extinction, FamilyRelation, FDAPhase, IndicesChanges, Indictment, IPO, JointVenture, ManMadeDisaster, Merger, MilitaryAction, MovieRelease, MusicAlbumRelease, NaturalDisaster, PatentFiling, PatentIssuance, PersonAttributes, PersonCareer, PersonCommunication, PersonEducation, PersonEmailAddress, PersonLocation, PersonParty, PersonRelation, PersonTravel, PoliticalEndorsement, PoliticalRelationship, PollsResult, ProductIssues, ProductRecall, ProductRelease, Quotation, SecondaryIssuance, StockSplit, Trial, VotingResult

4.2.1   Entities

Click a link for a description of the entity type and its attributes.

Anniversary, City, Company, Continent, Country, Currency, Editor, EmailAddress, EntertainmentAwardEvent, Facility, FaxNumber, Holiday, IndustryTerm, Journalist, MarketIndex, MedicalCondition, MedicalTreatment, Movie, MusicAlbum, MusicGroup, NaturalFeature, OperatingSystem, Organization, Person, PharmaceuticalDrug, PhoneNumber, PoliticalEvent, Position, Product, ProgrammingLanguage, ProvinceOrState, PublishedMedium, RadioProgram, RadioStation, Region, SportsEvent, SportsGame, SportsLeague, Technology, TVShow, TVStation, URL

Note: It is important to note that most attributes are optional; a tag can be extracted with some but not all of its attributes.

For a conceptual explanation, see Entity Markup Tags.

4.2.1.1 Anniversary

em/e/Anniversary	
Definition	This tag is generated when Open Calais identifies a mention or indication of an anniversary.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the anniversary. When there are multiple mentions, the most complete mention is used.

Example 1

Input Text	Spain marked the first anniversary of the train bombings in Madrid that took 200 lives.
Extracted Entity	first anniversary of the train bombings in Madrid

Example 2

Input Text	The office of Serbian President Boris Tadic today said he will attend the 10th anniversary of the massacre of Bosnian Muslims by Serbian troops.
Extracted Entity	10th anniversary of the massacre of Bosnian Muslims by Serbian troops.

4.2.1.2   City

em/e/City	
Definition	The name of a city or other municipality. Districts or neighborhoods within a city (e.g. Soho, London) are not extracted.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
Related Tag	er/Geo/City (Disambiguation Tag)

Example1

Input Text	Finally, negotiations are expected to go through Friday night and into early Saturday morning in the Hague over a treaty to reduce global warming.
Extracted Entity	the Hague

Example2

Input Text	Three Russian soldiers died and six were wounded when a Russian military truck blew up a rebel mine near the southern village of Alkhan-Yurt.
Extracted Entity	Alkhan-Yurt

4.2.1.3   Company

em/e/Company	
Definition	A full or partial company name. The term "Company" refers to any business organization, including newspapers, media companies, law firms, etc.
Attributes	**confidencelevel**: A confidence score on a scale of 0 to 1. The value represents the probability that the extracted entity is indeed of the assigned type.
	For example:
	• The confidence level score for an em/e/person tag indicates the probability that the extracted person is indeed a person.
	• The confidence level score for an em/e/company tag indicates the probability that the extracted company is indeed a company.
	• The confidence level score for an em/e/pharmaceuticalDrug tag indicates the probability that the extracted pharmaceutical drug is indeed a pharmaceutical drug.
	The higher the value, the higher the probability.
	The consuming application can use this score to achieve higher accuracy results by ignoring instances with confidence scores below a specified level. Note that boosting Precision in this manner is at the expense of Recall.
	Every em/e tag that displays this attribute also has a corresponding Confidence tag.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	The confidencelevel determines if the forenduserdisplay value is true or false.
	inlineric: The company Reuters Instrument Code (RIC), if it is mentioned in the text.
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
	nationality: The nationality of the extracted Person, Company, or Organization, if it is mentioned in the text.
Related Tags	Confidence, Relevance, er/Company (Disambiguation Tag)

Example1

Input Text	Zenith National Insurance Corp. (NYSE:ZNT) reported net income of $20.8 million for the fourth quarter of 2003.
Extracted Entity	Zenith National Insurance Corp.

Example2

Input Text	We increased our ownership in Advent Capital (Holdings) PLC ("Advent Capital") in the third quarter ... was reduced by approximately $0.9 million after tax for our share of adverse development recorded by Advent Capital in the third quarter of 2003.
Extracted Entity	Advent Capital (Holdings) PLC

4.2.1.4   Continent

em/e/Continent	
Definition	A continent.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The entity name, based on a mention in the text and mapped to a value from an internal reference list.
Related Tag	er/Continent (Disambiguation Tag)

Example

Input Text	European Union countries now invest more in South America than does the United States.
Extracted Entity	South America

4.2.1.5 Country

em/e/Country	
Definition	A country.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The entity name, based on a mention in the text and mapped to a value from an internal reference list.
Related Tag	er/Geo/Country (Disambiguation Tag)

Example1

Input Text	He also dealt with refugees in Cyprus after the 1974 Turkish invasion.
Extracted Entity	Cyprus

Example2

Input Text	He gained widespread praise for overseeing East Timor's three-year transition to independence after Indonesia withdrew in 1999.
Extracted Entity	East Timor Indonesia

4.2.1.6   Currency

em/e/Currency	
Definition	A reference to a currency.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The entity name, based on a mention in the text and mapped to a value from an internal reference list. **permid:** The Thomson Reuters unique ID (PermID) for this metadata tag type. Can be used when building a knowledge graph.
Related Tag	er/Currency (Disambiguation Tag)

Example1

Input Text	UOB's new notes, priced yesterday at 3.5%, also scored the tightest spread over Singapore dollar SOR, at 108bp, beating DBS Bank's S$250m 12-year non-call seven T2 in January last year which yielded a spread of 110bp.
Extracted Entity	Singapore Dollar

Example2

Input Text	UOB, rated Aa1/AA-/AA-, doubled down on the success of the Singapore dollar notes this morning with the announcement of a proposed euro benchmark five-year and/or US dollar benchmark three-year Reg S covered bond issue
Extracted Entity	Euro Singapore Dollar US Dollar

4.2.1.7   Editor

em/e/Editor	
Definition	The editor/s of the article, as mentioned in the Editor section (usually the last section of the document).
	Note: Open Calais will not generate an em/e/person tag for a person mentioned only in the Editor section.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	(Editing by Keith Weir and Sara Ledwith)
Extracted Entity	Keith Weir
	Sara Ledwith

4.2.1.8   EmailAddress

em/e/EmailAddress	
Definition	An email address.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	For information contact: info@softwareshelf.co.uk.
Extracted Entity	info@softwareshelf.co.uk

4.2.1.9   EntertainmentAwardEvent

em/e/EntertainmentAwardEvent	
Definition	An entertainment-related event or award, such as a film or music festival, or an award related to the entertainment industry.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example 1

Input Text	This week at the Venice Film Festival, a woman pushed past guards and tried to hug Pitt before being pulled away.
Extracted Entity	Venice Film Festival

Example 2

Input Text	Ledger, 28, and Williams, 27 received Oscar nominations for their performances in "Brokeback Mountain" directed by Ang Lee.
Extracted Entity	Oscar

4.2.1.10   Facility

em/e/Facility	
Definition	The name of a man-made physical entity. • Things like courts, embassies, consulates, radio stations, and TV stations are assigned a Facility tag only if the text string contains one or more explicit location or building indicator, such as "the Parliament building," "The crowd outside the Stanislaus County Courthouse," etc. • Places like parks, borders, settlements, farms, parking lots, etc., are not assigned a Facility tag. • Theme parks and amusement parks are assigned a Facility tag. • Universities and hospitals are assigned a Facility tag.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example 1

Input Text	A walking group for seniors, sponsored by Suburban Hospital, meets at the Montgomery Mall in Bethesda.
Extracted Entity	Montgomery Mall

Example 2

Input Text	Today's meeting took place at the Pratica di Mare military base, Europe's second largest, located 30 kilometers south of Rome.
Extracted Entity	Pratica di Mare military base

Example 3

Input Text	The launch took place at the Welsh High School hall and was attended by children whom Sister Aidan had delivered.
Extracted Entity	Welsh High School hall

Example 4

Input Text	Singapore said some of the suspects had received training at al-Qaida terrorist camps in Afghanistan.
Extracted Entity	al-Qaida terrorist camps

4.2.1.11 FaxNumber

em/e/FaxNumber	
Definition	A fax number, including the prefix and extension if they are present in the input text.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example 1

Input Text	Fax the completed form to +49 (0) 89 613 676 33
Extracted Entity	+49 (0) 89 613 676 33

Example 2

Input Text	Complete the survey and send by fax to 334|473-6503.
Extracted Entity	334|473-6503

4.2.1.12   Holiday

em/e/Holiday	
Definition	The name of a holiday, including the country or nationality if mentioned.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The entity name, based on a mention in the text and mapped to a value from an internal reference list.

Example 1

Input Text	Israel had closed off the West Bank and Gaza Strip on Monday at the beginning of the four-day Muslim **Feast of the Sacrifice** holiday.
Extracted Entity	Eid Al-Adha (The mention, Feast of the Sacrifice, was mapped to the official name of the holiday, Eid Al-Adha, as defined in the internal dataset.)

Example 2

Input Text	Parades will take place during the **Latvian Foundation Day**.
Extracted Entity	Latvia's Foundation Day

Example 3

Input Text	St. Patrick's Day in Ireland has become completely Americanized.
Extracted Entity	St. Patrick's Day in Ireland

4.2.1.13   IndustryTerm

em/e/IndustryTerm	
Definition	A text string related to a specific industry, segment, product family, or business.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example 1

Input Text	Avionics has expanded its capability to fulfill its customers' product needs, and now provides a complete line of avionics products.
Extracted Entity	avionics products

Example 2

Input Text	IBM Life Sciences brings together IBM resources, including research, services and e-business expertise, and high-performance computing to offer new solutions for the life sciences market, including e-health and pharmaceutical industries.
Extracted Entity	computing
	e-business
	e-health and pharmaceutical industries

4.2.1.14   Journalist

em/e/Journalist	
Definition	The author of the article, as mentioned in the Journalist section (usually the first section of the document, indicated by the word "by").
	Note: Open Calais will not generate an em/e/person tag for a person mentioned only in the Journalist section.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	By Nellie Andreeva and Kimberly Nordyke
Extracted Entity	Nellie Andreeva
	Kimberly Nordyke

4.2.1.15   MarketIndex

em/e/MarketIndex	
Definition	A stock market index.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The entity name, based on a mention in the text and mapped to a value from an internal reference list.
	permid: The Thomson Reuters unique ID (PermID) for this metadata tag type. Can be used when building a knowledge graph.
	ric: The Market Index Reuters Instrument Code (RIC), if it is mentioned in the text.

Example 1

Input Text	The surprise entry of ad major Publicis into the CAC 40.
Extracted Entity	CAC 40

Example 2

Input Text	Shinsei Bank and electronic materials firm Nitto Denko Corp. are leading candidates to join Tokyo's blue-chip Nikkei average after its annual overhaul.
Extracted Entity	Nikkei 225

4.2.1.16   MedicalCondition

em/e/MedicalCondition	
Definition	A medical condition such as a disease, disorder, or syndrome.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example 1

Input Text	After an operation for pancreatic cancer last year, Pavarotti had hoped to finish a world tour but he died on Thursday at the age of 71.
Extracted Entity	pancreatic cancer

Example 2

Input Text	First they tried nudging. Now companies are penalizing workers who have high health risks such as obesity and high blood pressure as insurance costs climb.
Extracted Entity	obesity high blood pressure

4.2.1.17   MedicalTreatment

em/e/MedicalTreatment	
Definition	A treatment, procedure, or therapy.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	Merck plans to conduct a Phase 3 trial of oral deforolimus in patients with metastatic soft-tissue and bone sarcomas following a favorable response to chemotherapy.
Extracted Entity	chemotherapy

4.2.1.18   Movie

em/e/Movie	
Definition	A movie or film.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	Tyler portrays Detective Wallis in the movie Death Sentence, which also stars Kevin Bacon as Nick Hume, a vigilante out for revenge.
Extracted Entity	Death Sentence

4.2.1.19   MusicAlbum

em/e/MusicAlbum	
Definition	A music album, CD, etc.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example	
Input Text	U.S. rapper 50 Cent released his new album, "Curtis," on MySpace Friday – days before the hotly anticipated disc is set to drop in stores.
Extracted Entity	Curtis

4.2.1.20   MusicGroup

em/e/MusicGroup	
Definition	A music group or band.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	On November 24, 1991, the world lost one of the greatest musical voices of all times with the passing of Freddie Mercury, the lead singer for the group Queen.
Extracted Entity	Queen

4.2.1.21   NaturalFeature

em/e/NaturalFeature	
Definition	A natural feature such as a river, sea, lake, or mountain, or the name of a geographical region.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	The Congolese refugees in Tanzania said they were forced to flee southwards, along the shores of Lake Tanganyika.
Extracted Entity	Lake Tanganyika

4.2.1.22   OperatingSystem

em/e/OperatingSystem	
Definition	An operating system.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	Jay reported yesterday that the Google Maps for mobile for Palm OS update would be available today.
Extracted Entity	Palm OS

4.2.1.23   Organization

em/e/Organization	
Definition	The name of an organization (governmental, military, or other organization).
	Note: A business organization that generates an em/e/Company tag (that is extracted as a company) will not generate an em/e/Organization tag.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name:** The organization name. **nationality**: The nationality of the extracted Person, Company, or Organization, if it is mentioned in the text. **organizationtype**: An organization classification. Possible values: sports, governmental military, governmental civilian, political party, N/A. **permid:** The Thomson Reuters unique ID (PermID) for this metadata tag type. Can be used when building a knowledge graph.
Related Tag	er/Organization (Disambiguation Tag)

Example 1

Input Text	Bond players will be focused on the BOJ's September Tankan corporate sentiment survey due out next Thursday.
Extracted Entity	Bank of Japan

Example 2

Input Text	Chief of the General Staff, Lt. Gen. Gabi Ashkenazi, met earlier this evening with the Commander of the Kenyan Military, General Jeremiah Mutinda.
Extracted Entity	Organization = Kenyan Military

4.2.1.24   Person

em/e/Person	
Definition	The name of a person. If the person is mentioned more than once in the document, the most complete mention is extracted.
Attributes	**commonname**: The commonly used person name.
	confidencelevel: A confidence score on a scale of 0 to 1. The value represents the probability that the extracted entity is indeed of the assigned type.
	For example:
	• The confidence level score for an em/e/person tag indicates the probability that the extracted person is indeed a person.
	• The confidence level score for an em/e/company tag indicates the probability that the extracted company is indeed a company.
	• The confidence level score for an em/e/pharmaceuticalDrug tag indicates the probability that the extracted pharmaceutical drug is indeed a pharmaceutical drug.
	The higher the value, the higher the probability.
	The consuming application can use this score to achieve higher accuracy results by ignoring instances with confidence scores below a specified level. Note that boosting Precision in this manner is at the expense of Recall.
	Every em/e tag that displays this attribute also has a corresponding Confidence tag.
	firstname: Most complete mention of the first name.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	lastname: Most complete mention of the last name.
	middlename: Most complete mention of the middle name.
	name: The name of the person as it appears in the text. If the person is mentioned more than once in the text, then the most complete mention is used.
	nationality: The nationality of the extracted Person, Company, or Organization, if it is mentioned in the text.
	permid: The Thomson Reuters unique ID (PermID) for this metadata tag type. Can be used when building a knowledge graph.
	persontype: A person classification. Possible values: sports, entertainment, political, economic, military, NA.
Related Tags	Confidence, Relevance, er/Person (Disambiguation Tag)

Example 1

Input Text	Prof. Tindall is survived by his son, Bruce Tindall of San Diego; daughter, Blair Tindall of Santa Monica, Calif., and one grandson.
Extracted Entity	George Brown Tindall
	Bruce Tindall
	Blair Tindall

Example 2

Input Text	The Palestinian Prime Minister Mahmoud Abbas, also known as Abu Mazen, visited Gaza yesterday. Abu Mazen told the reporters…
Extracted Entity	Person = Mahmoud Abbas (Abu Mazen)

4.2.1.25   PharmaceuticalDrug

This functionality is available upon subscription to Thomson Reuters Intelligent Tagging (TRIT). For more information please contact us at questions@opencalais.com.

em/e/PharmaceuticalDrug	
Definition	A pharmaceutical drug.
Attributes	**confidencelevel**: A confidence score on a scale of 0 to 1. The value represents the probability that the extracted entity is indeed of the assigned type. For example: • The confidence level score for an em/e/person tag indicates the probability that the extracted person is indeed a person. • The confidence level score for an em/e/company tag indicates the probability that the extracted company is indeed a company. • The confidence level score for an em/e/pharmaceuticalDrug tag indicates the probability that the extracted pharmaceutical drug is indeed a pharmaceutical drug. The higher the value, the higher the probability. The consuming application can use this score to achieve higher accuracy results by ignoring instances with confidence scores below a specified level. Note that boosting Precision in this manner is at the expense of Recall. Every em/e tag that displays this attribute also has a corresponding Confidence tag. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
Related Tags	Confidence, Relevance

Example

Input Text	Novartis AG's acquisition of a majority stake in Speedel Holding AG is a positive strategic move that will bring control over the blood pressure drug Tekturna and enhance its development pipeline, according to analysts.
Extracted Entity	Tekturna

4.2.1.26   PhoneNumber

em/e/PhoneNumber	
Definition	A phone number, complete with the prefix and extension if they are present in the input text.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	For further information call (203) 661-1926, ext. 6668.
Extracted Entity	**(203) 661-1926, ext. 6668**

4.2.1.27   PoliticalEvent

em/e/PoliticalEvent	
Definition	A political event such as an election, primary, vote, rally, demonstration, debate, speech, etc.
Attributes	**date**: An absolute date (in the YYYY-MM-DD format). This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date or a reference to a date (e.g. yesterday, next month), taken directly from the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **location**: Location of the event if it is mentioned in the text. **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used. **politicaleventtype**: A political event classification. For example, elections, primaries.

Example 1

Input Text	Conservative billionaire Silvio Berlusconi won Italy's general elections.
Extracted Entity	general elections

Example 2

Input Text	On January 20, 2009 the next president of the United States will be sworn in on the steps of the capitol.
Extracted Entity	the next president of the United States will be sworn in

Example 3

Input Text	Romney hunted for votes among Cuban-American Republicans at a rally in a Miami suburb, vowing to "never give in" to Cuban President Fidel Castro.
Extracted Entity	rally

4.2.1.28 Position

em/e/Position	
Definition	A position that a person held, holds, or will hold.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example 1

Input Text	Dave Rivello, special agent with Drug Enforcement Agency in Salisbury, said the drugs first were distributed mostly in the Salisbury area.
Extracted Entity	special agent

Example 2

Input Text	Former Paterson Mayor Martin G. Barnes pleaded guilty in July.
Extracted Entity	Mayor

Example 3

Input Text	A judge investigating suspicions that Total paid bribes to access markets has opened legal proceedings against one of the oil giant's Middle East directors.
Extracted Entity	Judge

4.2.1.29　Product

em/e/Product	
Definition	A product.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
	producttype: A product classification. Possible values: aircraft, car, drug, electronics, weapon, other.
Related Tag	er/Product (disambiguation tag)

Example 1

Input Text	Drug developer Mylan Laboratories Inc. said Monday the Food and Drug Administration approved its generic version of Ortho McNeil Pharmaceuticals' epilepsy treatment Topamax.
Extracted Entity	Topamax

Example 2

Input Text	Dealers cannot keep fuel-efficient cars like the Ford Focus in stock, and automakers cannot roll them off the assembly line fast enough to keep up with demand.
Extracted Entity	Ford Focus

4.2.1.30   ProgrammingLanguage

em/e/ProgrammingLanguage	
Definition	A programming language.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	Resolver One uses the Python programming language, which was first released in 1991.
Extracted Entity	Python

4.2.1.31 ProvinceOrState

em/e/ProvinceOrState	
Definition	A province, state, county, or other jurisdiction that is part of a country.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
Related Tag	er/Geo/ProvinceOrState (Disambiguation Tag)

Example 1

Input Text	The private placement will be units of flow-through common shares and common share purchase warrants to investors in Alberta and British Columbia.
Extracted Entity	Alberta British Columbia

Example 2

Input Text	The new general secretary, a member of the ethnic Tay minority from the northern province of Bac Can, said he will continue to fight against corruption among officials.
Extracted Entity	Bac Can

4.2.1.32   PublishedMedium

em/e/PublishedMedium	
Definition	A publication such as a newspaper, journal, or magazine.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example 1

Input Text	The company also declined to deny or confirm a report in The Boston Globe last month, which quoted a letter to State Street clients alerting them to a 42 percent decline this year in the State Street Limited Duration Bond Fund for institutional investors.
Extracted Entity	The Boston Globe

Example 2

Input Text	The Oakland Tribune reported Monday night the deal will pay Russell at least $31 million guaranteed.
Extracted Entity	The Oakland Tribune

Example 3

Input Text	The 400 richest Americans, including U.S. citizens now living abroad, as compiled by Forbes magazine.
Extracted Entity	Forbes

4.2.1.33   RadioProgram

em/e/RadioProgram	
Definition	A radio program.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	When Americans want to be a part of the national conversation, they turn to Talk of the Nation, NPR's mid-day news-talk show.
Extracted Entity	Talk of the Nation

4.2.1.34   RadioStation

em/e/RadioStation	
Definition	A radio station. This entity type primarily identifies U.S. radio stations and may identify non-U.S. radio stations as well.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	WBUR is a public radio station broadcasting from the campus of Boston University in Boston, Massachusetts.
Extracted Entity	WBUR

4.2.1.35 Region

em/e/Region	
Definition	A non-politically defined and not natural geographical region of the world (excluding continents). For example, the Middle East.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
Related Tag	er/Region (disambiguation tag)

Example

Input Text	The Company facilitates the movement of freight to and from anywhere in the world, with particular focus on business between the Far East and the United States.
Extracted Entity	Far East

4.2.1.36   SportsEvent

em/e/SportsEvent	
Definition	A sports event.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example 1

Input Text	The San Antonio Silver Stars may have fallen short of the ultimate goal of playing for the WNBA championship.
Extracted Entity	WNBA championship

Example 2

Input Text	The World Anti-Doping Agency yesterday welcomed the holding of an international summit on doping in cycling next month after this year's Tour de France was rocked by scandals.
Extracted Entity	Tour de France

4.2.1.37 SportsGame

em/e/SportsGame	
Definition	A reference to a sport such as Sumo, Swimming, Yachting, Football, etc.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	The Packers' Tracy White came out of the pile with the football for a touchdown.
Extracted Entity	Football

4.2.1.38   SportsLeague

em/e/SportsLeague	
Definition	The name of a sports league. Some sports leagues may also generate an em/e/organization tag.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	The team is being considered as one of six cities to host an expansion International Basketball League team for play in 2006.
Extracted Entity	International Basketball League

4.2.1.39   Technology

em/e/Technology	
Definition	Name or description of a type of technology.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example 1

Input Text	Vairex was founded in 1987 to develop advanced proprietary oil-free compressor technology.
Extracted Entity	oil-free compressor

Example 2

Input Text	The software's ability to stream over multiple formats has also been improved to now include MPEG-1, MPEG-2, WindowsMedia, RealVideo and QuickTime with MPEG-4 to be added soon.
Extracted Entity	MPEG-1 MPEG-2 MPEG-4

4.2.1.40   TVShow

em/e/TVShow	
Definition	A TV show.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	The actress, who stars in the new ABC series Pushing Daisies, fractured a few ribs, according to E!Online.com.
Extracted Entity	Pushing Daisies

4.2.1.41   TVStation

em/e/TVStation	
Definition	A TV station. This entity type primarily identifies U.S. TV stations and may identify non-U.S. TV stations as well.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	The Minnesota News Council narrowly upheld two complaints against KBJR-TV of Duluth and upheld one complaint against KSTP-TV of the Twin Cities on Thursday.
Extracted Entity	KBJR-TV KSTP-TV

4.2.1.42   URL

em/e/URL	
Definition	A URL or FTP address.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example 1

Input Text	http://www.macromedia.com/go/buymx2004
Extracted Entity	http://www.macromedia.com/go/buymx2004

Example 2

Input Text	http://reports.huginonline.com/915745/122470.pdf
Extracted Entity	http://reports.huginonline.com/915745/122470.pdf

Example 3

Input Text	ftp://ftp.demon.co.uk/pub/mirrors/garbo
Extracted Entity	ftp://ftp.demon.co.uk/pub/mirrors/garbo

4.2.2   Relations

Click a link for a description of the relation type and its attributes.

Acquisition, Alliance, AnalystEarningsEstimate, AnalystRecommendation, ArmedAttack, ArmsPurchaseSale, Arrest, Bankruptcy, BonusSharesIssuance, BusinessRelation, Buybacks, CandidatePosition, CompanyAccountingChange, CompanyAffiliates, CompanyCompetitor, CompanyCustomer, CompanyEarningsAnnouncement, CompanyEarningsGuidance, CompanyEmployeesNumber, CompanyExpansion, CompanyForceMajeure, CompanyFounded, CompanyInvestigation, CompanyInvestment, CompanyLaborIssues, CompanyLayoffs, CompanyLegalIssues, CompanyListingChange, CompanyLocation, CompanyMeeting, CompanyNameChange, CompanyProduct, CompanyReorganization, CompanyRestatement, CompanyTechnology, CompanyTicker, CompanyUsingProduct, ConferenceCall, ContactDetails, Conviction, CreditRating, Deal, DebtFinancing, DelayedFiling, DiplomaticRelations, Dividend, EmploymentChange, EmploymentRelation, EnvironmentalIssue, EquityFinancing, Extinction, FamilyRelation, FDAPhase, IndicesChanges, Indictment, IPO, JointVenture, ManMadeDisaster, Merger, MilitaryAction, MovieRelease, MusicAlbumRelease, NaturalDisaster, PatentFiling, PatentIssuance, PersonAttributes, PersonCareer, PersonCommunication, PersonEducation, PersonEmailAddress, PersonLocation, PersonParty, PersonRelation, PersonTravel, PoliticalEndorsement, PoliticalRelationship, PollsResult, ProductIssues, ProductRecall, ProductRelease, Quotation, SecondaryIssuance, StockSplit, Trial, VotingResult

Note: It is important to note that most attributes are optional; a tag can be extracted with some but not all of its attributes.

For a conceptual explanation, see Entity Markup Tags.

4.2.2.1   Acquisition

em/r/Acquisition	
Definition	A reference to an acquisition (past, present, or future) of shares in a company or of assets of another company.
Attributes	**company_acquirer:** Company that is acquiring the shares or assets.
	company_beingacquired: The company whose shares or assets are being acquired.
	confidencelevel: A confidence score on a scale of 0 to 1. The value represents the probability that the extracted relation is indeed of the assigned type.
	For example:
	• The confidence level score for an em/r/deal tag indicates the probability that the deal is indeed a deal between the companies specified by the Acquirer and Target attributes.
	• The confidence level score for an em/r/bankruptcy tag indicates the probability that text actually refers to a bankruptcy involving the company specified by the Company attribute.
	• The confidence level score for an em/r/IPO tag indicates the probability that the text actually refers to an IPO involving the company specified by the Company attribute.
	The higher the value, the higher the probability.
	The consuming application can use this score to achieve higher accuracy results by ignoring instances with confidence scores below a specified level. Note that boosting Precision in this manner is at the expense of Recall.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	status: The status, based on an indication in the text.

Example 1

Input Text	Ericsson Business Innovation, the innovation development company recently set up by Ericsson (NASDAQ:ERICY), is buying a 29 percent stake in Mediatude, a company specializing in targeted mobile marketing services over voice, SMS and WAP.
Extracted Relation	company_acquirer = Ericsson Business Innovation company_beingacquired = Mediatude status = announced

Example 2

Input Text	EMI Music Group said in September it had opened formal talks to buy Warner Music that would create a music firm second only to Universal Music and be home to a range of music from the Beatles and Frank Sinatra to Madonna and Norah Jones.
Extracted Relation	company_acquirer = The EMI Group company_beingacquired = Warner Music date = 2006-09-00 datestring = September status = planned

4.2.2.2 Alliance

em/r/Alliance	
Definition	A reference to an alliance agreement signed recently or in the past between two or more companies.
Attributes	**company:** Company that is a member of the alliance.
	Note: The company attribute appears two or more times in the em/r/alliance tag. Each company attribute indicates a different company that is a member of the alliance.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	status: The status, based on an indication in the text.

Example

Input Text	Digital Marketing Services, Inc. (DMS), the leading provider of online marketing research and a division of America Online Inc. (AOL), today announced an alliance with Netcentives Inc. (Nasdaq: NCNT)
Extracted Relation	company = Digital Marketing Services, Inc.
	company = Netcentives Inc.
	status = announced
	datestring = today
	date = 2000-01-31

4.2.2.3   AnalystEarningsEstimate

em/r/AnalystEarningsEstimate	
Definition	An estimation by external analysts regarding the future (i.e., not-yet-reported) financial results of a publicly-traded company.
Attributes	**company_rated**: Company that is the subject of the estimate, recommendation, or rating. **company_source**: The financial or other body publishing the estimate, rating, or recommendation. **financialmetric:** The reported metric, based on a mention in the text and mapped to one of the following values: - Revenues - Financial Results - EARNINGS - EBIT - EBITDA - FFO - Loss All but Revenues and Financial Results may be suffixed by: (BASIC), (BASIC_AND_DILUTED), or (DILUTED). **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **person_source**: The person who issued the estimate, rating, or recommendation. **quarter**: Time period to which the extracted relation refers. **year**: The relevant year.

Example 1

Input Text	Shares of Oplink Communications Inc., a provider of optical manufacturing, design and integration services, rocketed in afternoon trading on Friday after the company … Analysts polled by Thomson Financial expect earnings of 60 cents per share on $75.7 million in sales for 2007.
Extracted Relation	company_rated = Oplink Communications Inc. financialmetric = earnings per share (diluted) quarter = FY year = 2007

Example 2

Input Text	After casual dining chain California Pizza Kitchen Inc. boosted its second-quarter profit outlook on better-than-expected sales, ThinkEquity analyst Nicole M. Miller said the company is positioned to continue growing long-term with new menu items. Miller increased her second-quarter estimate a penny to 30 cents per share and reiterated her "Buy" rating on the stock.
Extracted Relation	company_source = ThinkEquity person_source = Nicole M. Miller company_rated = California Pizza Kitchen Inc. financialmetric = Earnings Per Share (Diluted) quarter = Q2

4.2.2.4   AnalystRecommendation

em/r/AnalystRecommendation	
Definition	A recommendation by an analyst (of an investment bank or similar organization) regarding shares of a publicly traded company. This tag type extracts recommendations regarding shares (equity) only. A recommendation regarding bonds or other credit instruments is not extracted by this tag type.
Attributes	**company_rated**: Company that is the subject of the estimate, recommendation, or rating. **company_source**: The financial or other body publishing the estimate, rating, or recommendation. **financialtrend**: The trend of the current recommendation with respect to the previous recommendation. Possible values: • Downgraded • Initiated (new recommendation as far as we know.) • Reiterated (same recommendation given again) • Upgraded **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **person_source**: The person who issued the estimate, rating, or recommendation. **price_new**: Target price quoted by the analyst. **price_old**: Previous target price, quoted by the same analyst. **rank_new**: New or current rank given to the company stock by the analyst. (For example, "Strong Buy," "Hold," etc.) **rank_old**: Previous company rank (as published by the same analyst).

Example 1

Input Text	RBC analyst Jordan Rohan upgraded aQuantive to "Outperform" from "Sector Perform" and raised the price target to $27 per share, from $25.
Extracted Relation	company_source = RBC person_source = Jordan Rohan company_rated = aQuantive financialtrend = upgraded rank_new = Outperform rank_old = Sector Perform price_new = $27 price_old = $25

em/r/AnalystRecommendation

Example 2

Input Text	Prudential analyst John McPeake initiated coverage of three software companies Tuesday, … McPeake also initiated coverage of Salesforce.com, which provides customer relationship management software and services via Web browsers, and Art Technology Group Inc., which makes e-commerce and marketing software.
Extracted Relation	company_source = Prudential person_source = John McPeake company_rated = Salesforce.com company_rated = Art Technology Group Inc. financialtrend = initiated

4.2.2.5   ArmedAttack

em/r/ArmedAttack	
Definition	An armed attack by a country or an organization, or an armed attack on a leader or political figure.
Attributes	**attacker**: The name of the attacker. (Could be the name of a group, a party, a person, etc.). **attacktype**: Type of attack. (For example: war, incursion, skirmish, attach, explosion, armed patrol, coup, etc.) **casualties**: Number and degree of casualties in the attack. **date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **locationstring:** Location of the armed attack, as mentioned in the text. **partyinvolved1**: Country, person, or group involved in the attack. **partyinvolved2:** Country, person, or group involved in the attack. **Note:** If it is not clear which side is the attacker and which side is the attacked, they will be extracted as partyinvolved1 and partyinvolved2 in random order. Otherwise, partyinvolved1 is the attacker, partyinvolved2 is the attacked, and the attacker attribute will have the same value as the partyinvolved1 attribute. **status:** The status, based on an indication in the text. **weaponused**: Weapons used to attack.

Example 1

Input Text	A roadside bomb killed three pilgrims walking to a major Shiite religious event south of Baghdad on Monday.
Extracted Relation	attacktype = Explosion weaponused = a roadside bomb status = confirmed casualties = killed three pilgrims date = 2012-03-28

Example 2

Input Text	Since the U.S.-led invasion in March 2003, 3,957 U.S. troops have been killed
Extracted Relation	attacker = U.S. attacktype = incursion status = confirmed casualties = 3,957 U.S. troops have been killed

em/r/ArmedAttack

Example 3

Input Text	Congolese Tutsi rebels and Mai Mai militia clashed on Monday in eastern Democratic Republic of Congo.
Extracted Relation	partyinvolved1 = Congolese Tutsi rebels partyinvolved2 = Mai Mai militia attacktype = skirmish status = confirmed

4.2.2.6   ArmsPurchaseSale

em/r/ArmsPurchaseSale	
Definition	A reference to an arms deal.
Attributes	**armsdescription**: Types, amounts, and models of weaponry and vehicles included in the deal.
	armsseller: The party selling the arms.
	countryarmspurchaser: Country which is purchasing the arms.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	moneyamount: The value of the deal.
	organizationarmspurchaser: The organization purchasing arms.
	personarmspurchaser: The person purchasing the arms.
	status: The status, based on an indication in the text.

Example 1

Input Text	A book by an FBI consultant on international terrorism says that Osama bin Laden's al-Qaeda terrorist network purchased 20-suitcase nuclear weapons for $30 million in 1998.
Extracted Relation	armsdescription = 20-suitcase nuclear weapons
	moneyamount = $30 million
	organizationarmspurchaser = al-Qaeda
	status = completed
	date = 1998-00-00

Example 2

Input Text	The Saudi arms deal stems from the American proposal to offer Gulf Arab states some $20 billion in weapons, including Joint Direct Attack Munition (JDAM) bomb kits for the Saudis.
Extracted Relation	armsdescription = weapons, including Joint Direct Attack Munition (JDAM) bomb kits
	moneyamount = $20 billion
	countryarmspurchaser = Saudi Arabia
	status = proposed

4.2.2.7   Arrest

em/r/Arrest	
Definition	A reference to the arrest of one or more people.
Attributes	**charge**: Phrase describing an offense/charge. In case of multiple charge attributes, the first charge attribute is the primary charge. **date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **othercharges**: Additional charges. **person**: The person who is the subject of this extracted relation.

Example 1

Input Text	Issac's brother Ramzi Issac, who was arrested along with the London suspect, was also charged with possessing false documents.
Extracted Relation	person = Ramzi Issac charge = possessing false documents

Example 2

Input Text	Also arrested Monday were his 33-year-old son Justin Zivojinovich and his wife, Michelle, 30.
Extracted Relation	person = Justin Zivojinovich date = 2009-03-02 datestring = Monday
Extracted Relation	person = Michelle date = 2009-03-02 datestring = Monday

4.2.2.8   Bankruptcy

em/r/Bankruptcy	
Definition	Extracts references to bankruptcy issues that involve companies: • Corporate insolvencies and bankruptcies, where a company has filed for bankruptcy or been declared insolvent. • Voluntary and involuntary insolvency proceedings including creditor protection actions such as Chapter 11 federal bankruptcy code proceedings in the U.S. and administration orders elsewhere as well as receiverships. • Company ceasing trading, bankruptcy court rulings and a company's emergence from a bankruptcy protection or administration process.
Attributes	**company**: The company that is the subject of the relation type. **confidencelevel**: A confidence score on a scale of 0 to 1. The value represents the probability that the extracted relation is indeed of the assigned type. For example: • The confidence level score for an em/r/deal tag indicates the probability that the deal is indeed a deal between the companies specified by the Acquirer and Target attributes. • The confidence level score for an em/r/bankruptcy tag indicates the probability that text actually refers to a bankruptcy involving the company specified by the Company attribute. • The confidence level score for an em/r/IPO tag indicates the probability that the text actually refers to an IPO involving the company specified by the Company attribute. The higher the value, the higher the probability. The consuming application can use this score to achieve higher accuracy results by ignoring instances with confidence scores below a specified level. Note that boosting Precision in this manner is at the expense of Recall. Note that this em/r/tag also has a corresponding Confidence tag. **date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **status:** The status, based on an indication in the text.
Related Tags	[Confidence Tag](#)

Example 1

Input Text	German holding company WCM Inc. said on Wednesday it had applied for insolvency.
Extracted Relation	company = WCM Inc. status = applied datestring = Wednesday date = 2009-03-04

Example 2

Input Text	DOV Pharma Inc. says it may file for bankruptcy.
Extracted Relation	company = Dov Pharma Inc. status = possible apply

em/r/Bankruptcy

Example 3

Input Text	Winn-Dixie ready to seek OK to exit bankruptcy.
Extracted Relation	company = Winn-Dixie status = expected to emerge

4.2.2.9   BonusSharesIssuance

em/r/BonusSharesIssuance	
Definition	A reference to bonus shares being issued (in the past, present, or future).
Attributes	**bonussharesratio**: The ratio of the number of shares held by the shareholders to the number of shares they will receive.
	company: The company that is the subject of the relation type.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	status: The status, based on an indication in the text.

Example

Input Text	The Capital Market Authority said Saudi British Bank, an affiliate of HSBC Holdings, could issue one bonus share for every two shares held, increasing the number of the bank's outstanding shares to 1.5 billion from 1 billion.
Extracted Relation	company = HSBC Holdings PLC
	bonussharesratio = 1:2
	status = rumored

4.2.2.10   BusinessRelation

em/r/BusinessRelation	
Definition	A reference to a business agreement signed recently or in the past, between two or more companies.
	Note: Agreements phrased as "alliances" or "partnerships" are extracted as an Alliance. A "weaker" contact is extracted as a BusinessRelation.
Attributes	**businessrelationtype**: Defines the business relationship, as mentioned in the text. For example: licensing agreement. **company:** Company that is involved in the business relation.
	Note: The company attribute appears multiple times in the em/r/BusinessRelation tag. Each company attribute indicates a different company that is involved in the business activity.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **status:** The status, based on an indication in the text.

Example 1

Input Text	SmartForce (Nasdaq:SMTF), the world's largest e-Learning company, today announced an agreement with DigitalMed, a subsidiary of Tenet Healthcare Corporation (NYSE:THC), to provide e-Learning for the health care industry.
Extracted Relation	company = SmartForce company = DigitalMed status = announced businessrelationtype = agreement datestring = today date = 2001-04-18

Example 2

Input Text	LookSmart Ltd. said it swung to a third-quarter profit from its year-ago loss as revenues climbed 70 percent ahead of the January end of its key licensing agreement with MSN.
Extracted Relation	company = LookSmart Ltd. company = MSN status = cancelled businessrelationtype = key licensing agreement

4.2.2.11   Buybacks

em/r/Buybacks	
Definition	Extracts references to a company's plan to repurchase its own securities in the open market (buyback events).
Attributes	**company**: The company that is the subject of the relation type. **date**: Date related to the buyback, or the announcement date (in the YYYY-MM-DD format). This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **status:** The status, based on an indication in the text.

Example 1

Input Text	EarthLink, Inc. (Nasdaq: ELNK), one of the nation's leading internet service providers (ISPs), today announced that its Board of Directors has expanded its share repurchase program, authorizing the purchase of an additional $100 million of its outstanding shares of common stock.
Extracted Relation	company = EarthLink, Inc. datestring = today date = 2007-02-11 status = announced

Example 2

Input Text	The program will take effect upon completion of the Commercial Capital Bancorp's stock repurchase program which was authorized in May 2004.
Extracted Relation	company = Commercial Capital Bancorp, Inc. datestring = May 2004 date = 2004-05-00 status = planned

4.2.2.12   CandidatePosition

em/r/CandidatePosition	
Definition	Extracts information about political election candidates, including mayoral candidates, for current or past elections and the office (position) in question. The position can be one that the candidate aspires to, or one currently or previously held.
Attributes	**candidate**: Name of the candidate. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **positionaspirational**: The position to which the person aspires. **positioncurrent**: Candidate's current position. **positionpast**: A candidate's past position.

Example

Input Text	Santorum, a former U.S. senator from Pennsylvania, and Ron Paul, a Texas congressman with libertarian views, garnered a larger percentage of the crucial conservative vote and are now the leading alternatives to front-runner Romney, the former Massachusetts governor, as the candidates prepare for the next voting contest in New Hampshire.
Extracted Relation	candidate = Rick Santorum positionpast = Senator positionaspirational = President
Extracted Relation	candidate = Ron Paul positioncurrent = congressman positionaspirational = President
Extracted Relation	candidate = Mitt Romney positionpast = governor positionaspirational = President

4.2.2.13   CompanyAccountingChange

em/r/CompanyAccountingChange	
Definition	Extracts references to a company's change in its accounting method or policy. For example, a changed fiscal year, a change in the depreciation policy, or new accounting standards.
Attributes	**accountingchangetype**: The changed accounting method or policy. Possible values: • FY (fiscal year) • Depreciation policy • Accounting standard. **company**: The company that is the subject of the relation type. **date**: Date of the announcement of the change, (in the YYYY-MM-DD format). This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **status:** The status, based on an indication in the text.

Example 1

Input Text	On June 2, 2004, Elizabeth Arden, Inc.'s Board of Directors approved a fiscal year-end change from January 31 to June 30.
Extracted Relation	company = Elizabeth Arden, Inc. accountingchangetype = FY date = 2004-06-02 datestring = June 2, 2004 status = announced

Example 2

Input Text	Lockheed Martin Corporation (NYSE: LMT) today announced that, effective January 1, 1999, it will adopt a new accounting standard pertaining to costs related to start-up activities.
Extracted Relation	company = Lockheed Martin Corporation accountingchangetype = accounting standard date = 2009-02-08 datestring = today status = announced

4.2.2.14   CompanyAffiliates

em/r/CompanyAffiliates	
Definition	Extracts references to a parent company or subsidiary relationship between two companies.
Attributes	**affiliaterelationtype**: Nature of the relationship between CompanyAffiliate and Company_Parent. For example: subsidiary, division, spinoff, child, parent. **company_affiliate**: Company that is a subsidiary or division of another company. **company_parent**: Company that is a parent of another company. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).

Example 1

Input Text	To learn more about White Electronic Designs' subsidiary Panelview Incorporated, please visit its website.
Extracted Relation	company_parent = White Electronic Designs company_affiliate = Panelview Incorporated affiliaterelationtype = subsidiary

Example 2

Input Text	NDS Group plc (Nasdaq: NNDS; EASDAQ), a News Corporation company and Philips Semiconductors, a division of Royal Philips Electronics (NYSE: PHG; AEX: PHI) announced an agreement.
Extracted Relation	company_parent = News Corporation Company company_affiliate = NDS Group plc affiliaterelationtype = child
Extracted Relation	company_parent = Royal Philips Electronics company_affiliate = Philips Semiconductors affiliaterelationtype = division

Example 3

Input Text	Altria announced that it would spin off its international division, Philip Morris International.
Extracted Relation	company_parent = Altria Group, Inc. company_affiliate = Philip Morris International affiliaterelationtype = spinoff

4.2.2.15   CompanyCompetitor

em/r/CompanyCompetitor	
Definition	A competitor of the company.
Attributes	**company**: The company that is the subject of the relation type.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).

Example

Input Text	Shares of EDS's closest competitors - Infosys Technologies (nasdaq:INFY - news - people), Computer Sciences Corporation (nyse: CSC - news - people) and Accenture (nyse: ACN - news - people) - were all down Tuesday as investors did the math.
Extracted Relation	company = EDS company = Infosys Technologies
Extracted Relation	company = EDS company = Computer Sciences Corporation
Extracted Relation	company = EDS company = Accenture

4.2.2.16   CompanyCustomer

em/r/CompanyCustomer	
Definition	A customer of the company.
Attributes	**company_customer**: Company that purchases/purchased products or services from the company_provider. **company_provider**: Company that sells the product or service. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **organization_customer**: Organization that purchases/purchased products or services from the company_provider.

Example

Input Text	Curtis's customers include mass merchandisers such as Bi-Way Stores, Wal-Mart and Family Dollar.
Extracted Relation	company_provider = Curtis company_customer = Bi-Way Stores
Extracted Relation	company_provider = Curtis company_customer = Wal-Mart
Extracted Relation	company_provider = Curtis company_customer = Family Dollar

4.2.2.17   CompanyEarningsAnnouncement

em/r/CompanyEarningsAnnouncement	
Definition	Extracts references to an official announcement, made by the publicly traded company itself, regarding its financial earnings results.
Attributes	**company**: The company that is the subject of the relation type. **financialmetric:** The reported metric, based on a mention in the text and mapped to one of the following values: • Revenues • Financial Results • EARNINGS • EBIT • EBITDA • FFO • Loss All but Revenues and Financial Results may be suffixed by: (BASIC), (BASIC_AND_DILUTED), or (DILUTED). **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **permid:** The Thomson Reuters unique ID (PermID) for this metadata tag type. Can be used when building a knowledge graph. **quarter**: Time period to which the extracted relation refers. **year**: The relevant year.

Example 1

Input Text	Announcing the publication itself of the financial results: Stein Mart, Inc. (Nasdaq: SMRT - News) today announced financial results for its first quarter.
Extracted Relation	company = Stein Mart, Inc. quarter = Q1 financialmetric = Financial Results

Example 2

Input Text	The Hershey Company (NYSE: HSY) today announced sales and earnings for the fourth quarter and year ended December 31, 2008. Consolidated net sales were $1,377,380,000 compared with $1,342,222,000 for the fourth quarter of 2007.
Extracted Relation	company = Hershey Food Corporation quarter = FY year = 2008 financialmetric = Earnings
Extracted Relation	company = Hershey Food Corporation quarter = Q4 year = 2008 financialmetric = Earnings

4.2.2.18   CompanyEarningsGuidance

em/r/CompanyEarningsGuidance	
Definition	Extracts references to projected earnings, made by the publicly traded company itself.
Attributes	**company**: The company that is the subject of the relation type. **financialmetric:** The reported metric, based on a mention in the text and mapped to one of the following values: • Revenues • Financial Results • EARNINGS • EBIT • EBITDA • FFO • Loss All but Revenues and Financial Results may be suffixed by: (BASIC), (BASIC_AND_DILUTED), or (DILUTED). **financialtrend**: Indicates whether projected earnings are higher or lower than the previous earnings announcement. Possible values: • Higher • Lower **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **quarter**: Time period to which the extracted relation refers. **year**: The relevant year.

Example

Input Text	The Hershey Company (NYSE: HSY) today announced sales and earnings for the fourth quarter and year ended December 31, 2008. "Hershey's strong fourth quarter results represent a solid end to the year" said David J. West, President and Chief Executive Officer. "For 2009, we expect net sales growth of 2-3 percent."
Extracted Relation	company = Hershey Food Corporation quarter = FY year = 2009 financialmetric = Revenues financialtrend = Higher

4.2.2.19   CompanyEmployeesNumber

em/r/CompanyEmployeesNumber	
Definition	The number of workers in a company.
Attributes	**company**: The company that is the subject of the relation type.
	employeesnumber: The number of people employed by the company.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	location: Location related to the extracted relation (address or other physical description).
	unit: The name of the department or unit.

Example 1

Input Text	Wachovia has 120,000 employees nationwide.
Extracted Relation	company = Wachovia Corp. employeesnumber = 120,000 location = nationwide

Example 2

Input Text	The Coca-Cola Co. says it is cutting information technology jobs in a bid to become more efficient. Company spokesman Dana Bolden would not say Thursday how many jobs will be eliminated, but said there will not be a major reduction in staff. The IT department has 1,100 employees.
Extracted Relation	company = Coca-Cola Co. employeesnumber = 1,100 unit = IT

Example 3

Input Text	Nokia is focused on wireless and wired telecommunications, with 128,445 employees in 120 countries.
Extracted Relation	company = Nokia Oyj employees = 128,445 location = in 120 countries

4.2.2.20   CompanyExpansion

em/r/CompanyExpansion	
Definition	A company entering a new market or creating a new unit.
Attributes	**company**: The company that is the subject of the relation type. **date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **expansiontype**: "New Market" or "New Unit." **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **location**: Location related to the extracted relation (address or other physical description). **status:** The status, based on an indication in the text.

Example 1

Input Text	Lenovo Company enters netbook market.
Extracted Relation	company = Lenovo Company expansiontype = New Market status = known

Example 2

Input Text	French car manufacturer Peugeot Citroen has announced plans to build its first factory in Russia.
Extracted Relation	company = Peugeot SA expansiontype = New Market location = Russia status = planned

Example 3

Input Text	DGSE Companies, Inc. Announces Creation of Superior Precious Metals, Inc.
Extracted Relation	company = Dallas Gold and Silver Exchange expansiontype = New Unit status = announced

4.2.2.21   CompanyForceMajeure

em/r/CompanyForceMajeure	
Definition	Unexpected external event that interrupts the regular activity of a company.
Attributes	**company**: The company that is the subject of the relation type.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forcemajeure: The cause of the disruption of normal activity.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	location: Location related to the extracted relation (address or other physical description).

Example 1

Input Text	MEMC Electronic Materials, Inc. announced that a construction incident caused by one of its electrical subcontractors working on the Pasadena, Texas, polysilicon facility expansion resulted in a power outage to the entire site, although the power was eventually restored later the same day.
Extracted Relation	company = MEMC Electronic Materials Inc. location = Pasadena, Texas, United States forcemajeure = construction incident

Example 2

Input Text	A fire broke out at Finnish oil refiner Neste Oil's 50,000 barrels per day Naantali refinery in western Finland on Saturday, delaying restart of the plant by several days, the plant spokesman said.
Extracted Relation	company = Neste Oil Oyg location = Finland forcemajeure = fire date = 2008-10-11 datestring = on Saturday

4.2.2.22   CompanyFounded

em/r/CompanyFounded	
Definition	The year in which the company was founded.
Attributes	**company**: The company that is the subject of the relation type. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **year:** Year in which the extracted relation took place.

Example 1

Input Text	V-Access (India) Pvt. Ltd. was formed in 1996. It is a 100% subsidiary of Vadain Beheer Group.
Extracted Relation	company = V-Access (India) Pvt. Ltd. year = 1996

Example 2

Input Text	Hindustan Aeronautics Limited (HAL) came into existence on 1st October 1964.
Extracted Relation	company = Hindustan Aeronautics Limited year = 1964

4.2.2.23   CompanyInvestigation

em/r/CompanyInvestigation	
Definition	Reference to a company being investigated (for accounting or other issues).
Attributes	**company_investigated**: Company under investigation.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	investigationtype: The type of investigation being conducted.
	Possible values:
	• Accounting
	• Others
	organization_regulator: The regulator involved in the investigation.
	person_investigated: The person who is under investigation.

Example

Input Text	EchoStar Communications Corp (DISH.O) *on Wednesday said the U.S. Securities and Exchange Commission has inquired into its recent record-keeping and internal controls.
Extracted Relation	company_investigated = EchoStar Communications Corp
	investigationtype = Others
	organization_regulator = Securities and Exchange Commission
	date = 2004-12-31
	datestring = Wednesday

4.2.2.24   CompanyInvestment

em/r/CompanyInvestment	
Definition	An investment in a company.
Attributes	**company**: Company in which the investment was made.
	company_investor: Company making the investment.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	status: The status, based on an indication in the text.

Example

Input Text	Health Web site Drkoop.com Inc. said on Tuesday it received $20 million in equity financing, a new management team and a reconfigured board.
Extracted Relation	company = Drkoop.com Inc.
	status = announced
	datestring = Tuesday
	date = 2000-08-22

4.2.2.25   CompanyLaborIssues

em/r/CompanyLaborIssues	
Definition	Production stoppages brought about by labor disagreements, negotiations, etc.
Attributes	**company**: The company that is the subject of the relation type.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	location: Location related to the extracted relation (address or other physical description).
	status: The status, based on an indication in the text.

Example 1

Input Text	Workers at Greek telecoms group OTE staged a 24-hour strike on Tuesday to protest against planned changes they say will affect job security, as well as the group's further privatization.
Extracted Relation	company = Hellenic Telecommunication organization location = Greece date = 2008-09-16 datestring = on Tuesday status = known

Example 2

Input Text	A strike at Chile's Escondida, the world's biggest copper mine, entered a third week on Monday with the union rejecting a new wage offer and the company inviting workers to negotiate individual contracts.
Extracted Relation	company = Escondida location = Chile date = 2008-08-21 datestring = on Monday status = continued

4.2.2.26   CompanyLayoffs

em/r/CompanyLayoffs	
Definition	An occurrence of a company reducing its workforce. Open Calais extracts this metadata type based on mentions in the text, of the company cutting the number of employees, jobs, positions, etc.
Attributes	**company**: The company that is the subject of the relation type. **date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **employeesnumber**: The number of employees being laid off by the company. **employeespercentage**: The percentage of employees (out of the total company workforce) being laid off by the company. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **status:** The status, based on an indication in the text.

Example 1

Input Text	General Motors Corp. announced today a reduction in workforce.
Extracted Relation	company = General Motors Corp. status = announced datestring = today

Example 2

Input Text	During 2008, Hewlett-Packard laid off 2 percent of its staff, or about 1,770 workers.
Extracted Relation	company = Hewlett-Packard. status = known date = 2008-00-00 datestring = During 2008 employeespercentage = 2% employeesnumber = 1,770

Example 3

Input Text	Door-to-door newspaper sales will be dropped, and more than 200 telemarketing jobs will be eliminated, The Los Angeles Times said.
Extracted Relation	company = The Los Angeles Times status = planned employeesnumber = over 200

4.2.2.27   CompanyLegalIssues

em/r/CompanyLegalIssues	
Definition	A reference to Corporate Litigation or a Class Action involving a company.
Attributes	**company_plaintiff**: Company that is a plaintiff in a legal issue.
	company_sued: Company that is being sued.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	lawsuitclass: Classification of a legal issue.
	Possible values:
	• Class Action
	• Lawsuit
	person_plaintiff: The person who is a plaintiff in a legal issue.
	status: The status, based on an indication in the text.
	sueddescription: Description of the entity being sued.

Example 1

Input Text	A Dutch court ruled against navigation systems company TomTom on Thursday in a patent infringement lawsuit against rival navigation device maker IBM from the United States.
Extracted Relation	company_sued = International Business Machines Corporation
	company_plaintiff = TomTom BV
	lawsuitclass = lawsuit
	datestring = Thursday
	date = 2009-03-05

Example 2

Input Text	Eastman Kodak Co., the world's top maker of photographic film, said a federal court wins summary judgment in favor of the company in a class action lawsuit brought on behalf of shareholders.
Extracted Relation	company_sued = Eastman Kodak Co.
	lawsuitclass = class action
	status = wins summary judgment

4.2.2.28   CompanyListingChange

em/r/CompanyListingChange	
Definition	A reference to a company entering or leaving any known stock-exchange.
Attributes	**changetype**: Classification of the change. Possible values: • Enters • Leaves **company**: The company that is the subject of the relation type. **date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **stockexchange**: The name of the relevant stock market. **ticker**: The company's ticker symbol if mentioned in the text.

Example 1

Input Text	Camtek Ltd. chose to delist its shares from the Tel Aviv Stock Exchange in 2003.
Extracted Relation	company = Camtek Ltd stockexchange = TASE changetype = leaves date = 2003-00-00 datestring = in 2003

Example 2

Input Text	Getty Images Inc. will move from Nasdaq to NYSE.
Extracted Relation	company = Getty Images Inc. stockexchange = NASDAQ changetype = leaves
Extracted Relation	company = Getty Images Inc. stockexchange TASE changetype = enters

Example 3

Input Text	Marconi's American Depository Receipts Ltd. will trade on Nasdaq under the ticker symbol MRCIY.
Extracted Relation	company = Marconi's American Depository Receipts Ltd. stockexchange = NASDAQ changetype = enters

4.2.2.29   CompanyLocation

em/r/CompanyLocation	
Definition	The location of a company's headquarters (city, state, and country, if mentioned in the text).
Attributes	**city**: City related to the extracted relation. **company**: The company that is the subject of the relation type. **companylocationtype**: Functional description of the company location. For example: center of operations, headquarters, branch office. **country**: Country related to the extracted relation. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **provinceorstate**: The ProvinceOrState, related to the extracted relation.

Example 1

Input Text	RC Networks, headquartered in San Diego, Calif., USA, is a privately held, venture-backed company.
Extracted Relation	company = RC Networks city = San Diego, California, United States provinceorstate = California, United States country = United States companylocationtype = headquarters

Example 2

Input Text	Ore.-based PixelWorks Inc.
Extracted Relation	company = PixelsWorks Inc. provinceorstate = Oregon, United States country = United States companylocationtype = N/A

4.2.2.30   CompanyMeeting

em/r/CompanyMeeting	
Definition	A meeting held by a company for its shareholders.
Attributes	**city**: City related to the extracted relation. **company**: The company that is the subject of the relation type. **companymeetingtype**: The nature of the company meeting, based on topics to be discussed. For example: Annual General Meeting (AGM), Extraordinary General Meeting (EGM), Shareholders Meeting, etc. **country**: Country related to the extracted relation. **date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **facility**: The facility related to the extracted relation. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **provinceorstate**: The ProvinceOrState, related to the extracted relation. **status:** The status, based on an indication in the text.

Example 1

Input Text	Separately, ABN Amro said it would hold a shareholder meeting Sept. 20 to debate the merits of the two bids.
Extracted Relation	company = ABN Amro companymeetingtype = Shareholders Meeting status = announced datestring = Sept. 20 date = 2009-09-20

Example 2

Input Text	Amedia Networks, Inc. (OTC Bulletin Board: AANI - News), a provider of next generation media gateways to distribute and manage ultra-broadband triple-play services in the home, today announced that its 2007 Annual meeting of Stockholders will be held on Tuesday, August 14, 2007, at 8:30 AM at the Sheraton Eatontown Hotel.
Extracted Relation	company = Amedia Networks, Inc. companymeetingtype = Shareholders Meeting status = announced date = 2007-08-14 datestring = Tuesday, August 14, 2007 facility = Sheraton Eatontown Hotel

4.2.2.31   CompanyNameChange

em/r/CompanyNameChange	
Definition	A reference to a company name change.
Attributes	**company_formername**: The name of the company before the name change.
	company_newname: The new name of the company.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	status: The status, based on an indication in the text.
	stockexchange: The name of the relevant stock market.
	ticker: The company's new or updated stock symbol if mentioned in the text.

Example

Input Text	Troubled drug development company SFBC International said on Monday it has changed its name to PharmaNet Development Group Inc. and will be traded on Nasdaq under the stock symbol "PDGI".
Extracted Relation	company_formername = SFBC International
	company_newname = PharmaNet Development Group Inc.
	status = announced
	datestring = Monday
	date = 2008-12-22
	ticker = PDGI
	stockexchange = Nasdaq

4.2.2.32   CompanyProduct

em/r/CompanyProduct	
Definition	Extracts a reference to a company and its product/s.
Attributes	**company**: The company that is the subject of the relation type.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	product: The product that is the subject of the extracted relation.
	producttype: Product classification.
	Possible values:
	• aircraft
	• car
	• drug
	• electronics
	• weapon
	• other

Example

Input Text	Boeing is introducing the 787 Dreamliner which includes new technologies to create better environmental performance for commercial jetliners.
Extracted Relation	company = The Boeing Co.
	product = 787 Dreamliner
	producttype = aircraft

4.2.2.33 CompanyReorganization

em/r/CompanyReorganization	
Definition	A reference to a corporate reorganization or restructuring.
Attributes	**company**: The company that is the subject of the relation type.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	status: The status, based on an indication in the text.

Example 1

Input Text	Badger Income Fund said it will delay a planned corporate reorganization.
Extracted Relation	company = Badger Income Fund status = delayed

Example 2

Input Text	Avis Budget Group, Inc. announced Thursday the restructuring of its Budget Truck Rental subsidiary.
Extracted Relation	company = Avis Budget Group, Inc. status = announced datestring = Thursday date = 2006-10-26

4.2.2.34   CompanyRestatement

em/r/CompanyRestatement	
Definition	A reference to a company restatement of its financial statement for a specific time period.
Attributes	**company**: The company that is the subject of the relation type.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	quarter: Time period to which the extracted relation refers.
	status: The status, based on an indication in the text.
	year: The relevant year.

Example 1

Input Text	AMIS Holdings, Inc. announced that the audit committee has completed its review, and as a result, the Company has decided to restate its consolidated financial statements for the first and second quarters of fiscal year ended December 31, 2007.
Extracted Relation	company = AMIS Holdings, Inc. quarter = Q1 year = 2007 status = announced
Extracted Relation	company = AMIS Holdings, Inc. quarter = Q2 year = 2007 status = announced

Example 2

Input Text	Columbia Laboratories, Inc. announced that it is planning to restate its financial statements for the fiscal years ended December 31, 2005 and December 31, 2006.
Extracted Relation	company = Columbia Laboratories, Inc quarter = FY year = 2005 status = planned
Extracted Relation	company = Columbia Laboratories, Inc quarter = FY year = 2006 status = planned

4.2.2.35   CompanyTechnology

em/r/CompanyTechnology	
Definition	A technology associated with a company.
Attributes	**company**: Company associated with the technology. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **technology**: The technology related to the company.

Example 1

Input Text	Their expertise means that the company is ideally suited to distributing RC Networks' SDSL solutions, which complement the existing offerings.
Extracted Relation	company = RC Networks technology = SDSL

Example 2

Input Text	Targeted Genetics Corporation develops gene therapy products for the treatment of acquired and inherited diseases.
Extracted Relation	company = Targeted Genetics Corporation technology = gene therapy

4.2.2.36   CompanyTicker

em/r/CompanyTicker	
Definition	Extracted instance of a company name followed by its stock or ticker symbol/s in parentheses.
Attributes	**company**: The company that is the subject of the relation type. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **stockexchange**: The name of the relevant stock market. **ticker**: The company's ticker symbol if mentioned in the text.

Example

Input Text	Baltimore Technologies (NASDAQ:BALT; London:BLM)
Extracted Relation	company = Baltimore Technologies ticker = BALT stockexchange = NASDAQ
Extracted Relation	company = Baltimore Technologies ticker = BLM stockexchange = LSE

4.2.2.37   CompanyUsingProduct

em/r/CompanyUsingProduct	
Definition	A relationship that includes a company or an organization and a product that they are using.
Attributes	**company_customer**: Company that purchases/purchased products or services from the company_provider. **company_provider**: Company that sells the product or service. **date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **organization_customer**: Organization that purchases/purchased products or services from the company_provider. **product**: The product that is the subject of the extracted relation. **producttype**: Product classification. Possible values: • aircraft • car • drug • electronics • weapon • other **status:** The status, based on an indication in the text.

Example 1

Input Text	BakBone Software(R) (TSX: BKB, OTC Pink Sheets: BKBO), a global provider of data protection software, is using the AppExchange on-demand platform.
Extracted Relation	company_customer = BakBone Software product = AppExchange producttype = other status = announced

Example 2

Input Text	It has been announced that the US Army will use Microsoft's IPTV Edition software.
Extracted Relation	organization_customer = United States Army company_provider = Microsoft Corporation product = IPTV Edition producttype = other

4.2.2.38   ConferenceCall

em/r/ConferenceCall	
Definition	A conference call held by a company (in the past, present, or future).
Attributes	**company**: The company that is the subject of the relation type.
	conferencecalltype: The nature of the conference call, based on the topics to be discussed.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	quarter: Time period to which the extracted relation refers.
	status: The status, based on an indication in the text.

Example 1

Input Text	BlackBerry maker Research In Motion Ltd. postponed a conference call with investors and analysts to discuss second-quarter earnings, but gave no reason for the move.
Extracted Relation	company = Research In Motion Ltd. status = postponed conferencecalltype = earnings quarter = Q2

Example 2

Input Text	Raiffeisen Group and Bank Vontobel to hold joint conference call.
Extracted Relation	company = Raiffeisen Group status = planned
Extracted Relation	company = Bank Vontobel status = planned

4.2.2.39 ContactDetails

em/r/ContactDetails	
Definition	The contact details of a company or person.
	Note: Entities of type Person, Company, City, Country and ProvinceOrState, which appear only within a ContactDetails instance, will not generate em/e/Person, Company, City, Country, or Provinceorstate tags.
Attributes	**address**: The contact address. **companycontactentity**: The name of the company whose contact details are provided. **emailaddress1**: The contact e-mail address. **emailaddress2**: A second contact e-mail address. **fax**: The fax number of the contact. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **messenger**: The contact's Instant Messaging name. **personcontactentity1**: The first person whose contact details are provided. **personcontactentity2**: The second person whose contact details are provided. **telephone1**: The contact phone number. **telephone2**: A second contact phone number. **url**: Website of the contact.

Example

Input Text	For more information, see www.soldiersangels.org or call 615-676-0239. Tax ID# 20-0583415 CFC# 25131.
Extracted Relation	telephone1 = 615-676-0239 url = www.soldiersangels.org

4.2.2.40   Conviction

em/r/Conviction	
Definition	Reference to the conviction of one or more people.
Attributes	**charge**: Phrase describing an offense/charge. In case of multiple charge attributes, the first charge attribute is the primary charge. **date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **othercharges**: Additional charges. **person**: The person who is the subject of this extracted relation.

Example 1

Input Text	Mr. al-Sirri himself was sentenced to death in absentia in an Egyptian court in 1994 for his role in an assassination attempt on the Egyptian prime minister.
Extracted Relation	person = al-Sirri charge = assassination attempt on the Egyptian prime minister datestring = in1994 date = 1994-00-00

Example 2

Input Text	Guilherme "Will" Barbosa, 21, of Brockton, Mass., faces a maximum sentence of six years in prison when he is sentenced in Columbia County Court. He was found guilty of two aggravated assault counts, as well as lesser charges of simple assault and reckless endangerment.
Extracted Relation	person = Guilherme "Will" Barbosa charge = aggravated assault charge = simple assault charge = reckless endangerment

4.2.2.41   CreditRating

em/r/CreditRating	
Definition	A recent rating by a rating agency of a company, financial institution, or country.
Attributes	**company_rated**: Company that is the subject of the estimate, recommendation, or rating.
	company_source: The financial or other body publishing the estimate, rating, or recommendation.
	country_rated: Country which is being rated.
	financialtrend: The trend of the current rating with respect to the previous rating.
	Possible values:
	• Affirms
	• Assigns
	• Changes
	• Cuts
	• Expects to change
	• Puts
	• Raises
	• Rates
	• Removes
	• Says
	• Withdraws
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	organization_rated: The organization that is being rated.
	rank_new: New or current ranking of the company.
	rank_old: Previous ranking of the company.

Example 1

Input Text	HONG KONG (Standard & Poor's) Sept. 27, 2005--Standard & Poor's Ratings Services said today it had assigned its 'BB' long-term corporate credit rating to Fosun International Ltd. (Fosun).
Extracted Relation	company_source = Standard and Poor's Rating Services
	company_rated = Fosun International Ltd.
	financialtrend = Assigns
	rank_new = BB

em/r/CreditRating

Example 2

Input Text	This includes Fitch's 'A-' long-term and senior debt ratings and 'BBB+' preferred stock rating on RenRe Ltd., as well as its 'A+' insurer financial strength rating on lead operating subsidiary.
Extracted Relation	company_source = Fitch Ratings company_rated = RenRe Ltd. rank_new = A-
Extracted Relation	company_source = Fitch Ratings company_rated = RenRe Ltd. rank_new = BBB+
Extracted Relation	company_source = Fitch Ratings company_rated = RenRe Ltd. rank_new = A+

Example 3

Input Text	Fitch Ratings-New York-September 23, 2005: Fitch has upgraded UNOVA, Inc.'s (UNOVA) issuer default rating to 'BB-' from 'B-'
Extracted Relation	company_source = Fitch Ratings company_rated = UNOVA Inc. financialtrend = Raises rank_new = BB- rank_old = B-

4.2.2.42   Deal

This metadata type is available upon subscription to Thomson Reuters Intelligent Tagging (TRIT). For more information please contact us at questions@opencalais.com.

em/r/Deal	
Definition	A deal between two companies.
	The tags that describe a particular deal (one or more instance tags, em/r/deal, er/deal, confidence tag) are excluded from the output if Open Calais is not able to find the deal in the Thomson Reuters Deals Authority dataset.
	Note: The same text may also be identified by Open Calais as a Merger or an Acquisition. Open Calais generates em/r/merger and em/r/acquisition tags for identified Mergers and Acquisitions without attempting to locate them in the Thomson Reuters dataset.
Attributes	**acquirer**: The purchasing company.
	confidencelevel: A confidence score on a scale of 0 to 1. The value represents the probability that the extracted relation is indeed of the assigned type.
	For example:
	• The confidence level score for an em/r/deal tag indicates the probability that the deal is indeed a deal between the companies specified by the Acquirer and Target attributes.
	• The confidence level score for an em/r/bankruptcy tag indicates the probability that text actually refers to a bankruptcy involving the company specified by the Company attribute.
	• The confidence level score for an em/r/IPO tag indicates the probability that the text actually refers to an IPO involving the company specified by the Company attribute.
	The higher the value, the higher the probability.
	The consuming application can use this score to achieve higher accuracy results by ignoring instances with confidence scores below a specified level. Note that boosting Precision in this manner is at the expense of Recall.
	Note that this em/r/tag also has a corresponding Confidence tag.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	permid: The Thomson Reuters unique ID (PermID) for this metadata tag type. Can be used when building a knowledge graph.
	target: Company being acquired.
Related Tags	er/Deal (Disambiguation Tag), Confidence Tag, em/e/company

Example	
Input Text	Ericsson Business Innovation, the innovation development company recently set up by Ericsson (NASDAQ:ERICY), is buying a 29 per cent stake in Mediatude, a company specializing in targeted mobile marketing services over voice, SMS and WAP.
Extracted Relation	acquirer = Ericsson Business Innovation
	target = Mediatude

4.2.2.43   DebtFinancing

em/r/DebtFinancing	
Definition	A mention of companies seeking to finance their debt.
Attributes	**company**: Company raising money through debt financing.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	debtaction: The action taken. Extracted based on the context.
	debttype: The type of debt expressed as a financing instrument. For example, shares, bonds, notes, etc.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	status: The status, based on an indication in the text.

Example 1

Input Text	Verizon entered into a $7.55 billion 364-day credit facility.
Extracted Relation	company = Verizon Communications Inc. status = announced debttype = credit facility debtaction = entered

Example 2

Input Text	Time Warner Cable Inc. ("TWC"), a subsidiary of Time Warner Inc. completed its offering of $2.0 billion in aggregate principal amount of senior unsecured notes on 12th January 2008.
Extracted Relation	company = Time Warner Cable Inc. status = announced debttype = senior unsecured notes debtaction = completed date = 2008-01-12 datestring = 12th January 2008

4.2.2.44   DelayedFiling

em/r/DelayedFiling	
Definition	A mention of a delay in the filing of financial reports by a company.
Attributes	**company**: The company that is the subject of the relation type. **filingtype**: Name/s of the report/s filed, based on a mention in the text and mapped to one of the following values: • Form 10-K (for Form 10-K or Form 10-KSB) • Form 20-F • Form 10-Q (for Form 10-Q, 10-QSB, or other quarterly report) • Annual report • Interim report If the mention in the text cannot be mapped to one of the above values, the mention itself is used as the attribute value. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **quarter**: Time period to which the extracted relation refers. **status:** The status, based on an indication in the text. **year**: The relevant year.

Example

Input Text	King Pharmaceuticals, Inc. filed a Form 12b-25 today providing for a 15-calendar-day extension for submitting its Form 10-K for the year ended December 31, 2002, to the Securities and Exchange Commission ("SEC").
Extracted Relation	company = King Pharmaceuticals, Inc. filingtype = Form 10-K quarter = FY year = 2002 status = announced

4.2.2.45   DiplomaticRelations

em/r/DiplomaticRelations	
Definition	Information about diplomatic relations between diplomatic entities (e.g. countries, governments, minority groups, regions, international organizations, a spokesperson for a group, etc.).
Attributes	**date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **diplomaticaction**: Diplomatic action taken. **diplomaticentity1**: An organization or person relevant to the diplomatic action. **diplomaticentity2**: An organization or person relevant to the diplomatic action. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **location**: Location related to the extracted relation (address or other physical description).

Example

Input Text	Poland and the Baltic states came together on Saturday to issue a harsh condemnation of Russia for what they described as aggression in Georgia and urged NATO and the European Union to oppose it.
Extracted Relation	diplomaticentity1 = Poland diplomaticaction = condemnation diplomaticentity2 = Russia datestring = on Saturday date = 2008-08-09
Extracted Relation	diplomaticentity1 = Baltic States diplomaticaction = condemnation diplomaticentity2 = Russia datestring = on Saturday date = 2008-08-09

4.2.2.46   Dividend

em/r/Dividend	
Definition	Information about dividends paid to shareholders by a company.
Attributes	**company**: The company that is the subject of the relation type. **dividendtype**: The type of dividend. Possible values: • cash • shares **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **quarter**: Time period to which the extracted relation refers. **status:** The status, based on an indication in the text. **year**: The relevant year.

Example 1

Input Text	South African short-term insurer Santam reported annual results that beat expectations on Wednesday and said it will pay a special cash dividend, sending its stock nearly seven percent higher.
Extracted Relation	company = Santam status = announced dividendtype = cash

Example 2

Input Text	"We have a good level of yearly free cash flow, about $800 million for the entire Philippine Long Distance Telephone Co. group. Most likely we are enhancing our dividend payout to 70 percent of core earnings in 2007," he added.
Extracted Relation	company = Philippine Long Distance Telephone Co. status = planned quarter = FY year = 2007

4.2.2.47   EmploymentChange

em/r/EmploymentChange	
Definition	An announcement about a recent appointment, retirement, departure (including dismissal), or promotion of one or more officers in a company or organization.
Attributes	**changetype**: Classification of the change. Possible values: • Enters • Leaves • Retired **company**: The company that is the subject of the relation type. **date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **employmentchangestatus**: Status of the change in employment. Possible values: • announced • planned **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **organization**: The organization that is a participant in the extracted relation. **person**: The person who is the subject of this extracted relation. **position**: Position that the person held, holds, or will hold.

Example 1

Input Text	Electronic component developer Panja Inc. said it has appointed Berry Cash as chairman of the company.
Extracted Relation	company = Panja Inc. person = Berry Cash position = chairman changetype = enters employmentchangestatus = announced

em/r/EmploymentChange

Example 2

Input Text	Bank One Corp. said that president Verne Istock will retire Sept. 30.
Extracted Relation	company = Bank One Corp. person = Verne Istock position = president changetype = leaves datestring = Sept. 30 date = 2001-09-30

4.2.2.48   EmploymentRelation

em/r/EmploymentRelation	
Definition	Extracts a reference to a person's employment by another person.
Attributes	**date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **person_employee**: The person who is the employee. **person_employer**: The person who is the employer. **position**: Position that the person held, holds, or will hold. **status:** The status, based on an indication in the text.

Example

Input Text	Jan Scheck ended up being hired by Mr. E.J. Ridings at Trump Mortgage as national sales manager, but left the firm this year after his one-year contract was up.
Extracted Relation	person_employer = E.J. Ridings person_employee = Jan Scheck position = national sales manager status = former

4.2.2.49   EnvironmentalIssue

em/r/EnvironmentalIssue	
Definition	An issue that has an effect on the environment.
Attributes	**date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	effect: The impact of the environmental issue.
	environmentalissue: The environmental issue mentioned in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	location: Location related to the extracted relation (address or other physical description).

Example 1

Input Text	Each year we lose many coconut trees on our shore lines because of tidal erosion from sea level rise.
Extracted Relation	environmentalissue = sea level rise

Example 2

Input Text	Overfishing over historical times is largely responsible for the recent collapse of coastal ecosystems around North America.
Extracted Relation	environmentalissue = Overfishing
	location = around North America

4.2.2.50   EquityFinancing

em/r/EquityFinancing	
Definition	A secondary issuance of securities or a new issuance of other stocks.
Attributes	**company**: The company that is the subject of the relation type.
	currency: Currency in which the issuance is offered.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	status: The status, based on an indication in the text.

Example 1

Input Text	Last week SandRidge Energy, Inc. Announced Pricing of Public Offering of 15.2 Million Shares of Common Stock, at the price of AUD0.4344 per share, totaling to a purchase price of AUD198,000.
Extracted Relation	company = SandRidge Energy, Inc.
	status = Known
	datestring = Last week

Example 2

Input Text	Inversiones Copper Holding Chile Limitada will issue 1,993,750 common shares at a deemed price of $0.075 per share in satisfaction of the purchase price of $159,500.
Extracted Relation	company = Inversiones Copper Holding Chile Limitada
	status = planned

4.2.2.51   Extinction

em/r/Extinction	
Definition	Extracted mention/s of an extinct or endangered animal or plant.
Attributes	**date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **location**: Location related to the extracted relation (address or other physical description). **species**: The affected plant or animal species.

Example 1

Input Text	The Chinese paddle fish and the dog-eating catfish in Southeast Asia are on the brink of extinction because of pollution, overfishing and dam building.
Extracted Relation	species = The Chinese paddle fish location = Southeast Asia
Extracted Relation	species = dog-eating catfish location = Southeast Asia

Example 2

Input Text	In 1967, after years of over-hunting and habitat loss, the gator was listed as an endangered species.
Extracted Relation	species = the gator date = 1967-00-00 datestring = 1967

4.2.2.52   FamilyRelation

em/r/FamilyRelation	
Definition	A family relationship between two people.
Attributes	**familyrelationtype**: Describes the relationship of the person_relative to the person. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **person**: The person who is the subject of this extracted relation. **person_relative**: The person who is the family relative.

Example 1

Input Text	Gusti Made Mudita's daughter, Ni Gusti Ayu Kartini, 34, was missing in the explosion.
Extracted Relation	person = Gusti Made Mudita person_relative = Ni Gusti Ayu Kartini familyrelationtype = daughter

Example 2

Input Text	Mahathir Mohamad was to be accompanied by his wife Siti Hasmah Mohamad Ali.
Extracted Relation	person = Mahathir Mohamad person_relative = Siti Hasmah Mohamad Ali familyrelationtype = wife

4.2.2.53   FDAPhase

em/r/FDAPhase	
Definition	The stage of a drug in the US Food and Drug Administration's approval process.
Attributes	**company**: Name of the drug manufacturer. **fdastage**: The stage of the drug in the FDA approval process. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **product**: The product that is the drug in question. **producttype**: Product classification. Possible values: • aircraft • car • drug • electronics • weapon • other **status:** The status, based on an indication in the text.

Example

Input Text	Provectus Pharmaceuticals, Inc. (OTC Bulletin Board: PVCT) announced today it has filed an Investigational New Drug (IND) application for Provecta(TM), an advanced drug therapy designed to treat breast, liver, prostate and other potentially deadly cancers."
Extracted Relation	company = Provectus Pharmaceuticals, Inc. product = Provecta(TM) fdastage = INDA status = NA producttype = drug

4.2.2.54   IndicesChanges

em/r/IndicesChanges	
Definition	References to changes in financial indices, e.g. companies added to or removed from one of the indices. IndicesChanges are extracted only if a company is removed from an index for lack of presentation, and not because of a merger or acquisition.
Attributes	**changetype**: Classification of the change. Possible values: • Enters • Leaves **company**: The company that is the subject of the relation type. **date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **marketindex**: The index in which a change was made.

Example 1

Input Text	Tate & Lyle to replace Tomkins in UK's FTSE 100.
Extracted Relation	marketindex = FTSE 100 company = Tate & Lyle changetype = enters
Extracted Relation	marketindex = FTSE 100 company = Tomkins changetype = leaves

Example 2

Input Text	Euronext Amsterdam has added shipping group P&O Nedlloyd (NLYN.AS) and staffing company Vedior (VDOR.AS) to its main AEX Index in an annual reshuffle, the bourse operator said on Wednesday.
Extracted Relation	marketindex = AEX25 company = P&O Nedlloyd changetype = enters datestring = Wednesday
Extracted Relation	marketindex = AEX25 company = Vedior changetype = enters datestring = Wednesday

em/r/IndicesChanges

Example 3

Input Text	Just witness the effect on troubled telecom WorldCom on the day after the research firm said it would boot the troubled telecom company off the S&P 500 index.
Extracted Relation	marketindex = S&P 500 company = WorldCom changetype = leaves

4.2.2.55 Indictment

em/r/Indictment	
Definition	An indictment of a person or group of people (charged with criminal offenses).
Attributes	**charge**: Phrase describing an offense/charge. In case of multiple charge attributes, the first charge attribute is the primary charge. **date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **othercharges**: Additional charges. **person**: The person who is the subject of this extracted relation.

Example 1

Input Text	Ibrahim Muktar Said was charged Sunday night in connection with the failed Hackney bus bombing.
Extracted Relation	person = Ibrahim Muktar Said charge = failed Hackney bus bombing datestring = Sunday night date = 2009-03-08

Example 2

Input Text	Jermaine Wallace, 25, of Salisbury, and Natasha Adams, 25, of Bowie, were charged with conspiracy to distribute cocaine and money laundering.
Extracted Relation	person = Jermaine Wallace charge = conspiracy to distribute cocaine charge = money laundering
Extracted Relation	person = Natasha Adams charge = conspiracy to distribute cocaine charge = money laundering

4.2.2.56 IPO

em/r/IPO	
Definition	An initial public offering (IPO) of stock by a company.
Attributes	**company**: The company that is the subject of the relation type.
	confidencelevel: A confidence score on a scale of 0 to 1. The value represents the probability that the extracted relation is indeed of the assigned type.
	For example:
	• The confidence level score for an em/r/deal tag indicates the probability that the deal is indeed a deal between the companies specified by the Acquirer and Target attributes.
	• The confidence level score for an em/r/bankruptcy tag indicates the probability that text actually refers to a bankruptcy involving the company specified by the Company attribute.
	• The confidence level score for an em/r/IPO tag indicates the probability that the text actually refers to an IPO involving the company specified by the Company attribute.
	The higher the value, the higher the probability.
	The consuming application can use this score to achieve higher accuracy results by ignoring instances with confidence scores below a specified level. Note that boosting Precision in this manner is at the expense of Recall.
	Note that this em/r/tag also has a corresponding Confidence tag.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	permid: The Thomson Reuters unique ID (PermID) for this metadata tag type. Can be used when building a knowledge graph.
	status: The status, based on an indication in the text.
Related Tag	Confidence Tag

Example 1

Input Text	SM Investments Corp. has set the price of the Philippines' largest initial public offering at 250 pesos ($4.6) per share to raise a total 28.75 billion pesos ($530 million), traders said on Friday.
Extracted Relation	company = SM Investments Corp. status = announced datestring = Friday date = 2007-07-02

em/r/IPO

Example 2

Input Text	Science Applications International Corp. on Monday set its planned initial public offering at 75 million shares at an estimated price of $13 to $15 each.
Extracted Relation	company = Science Applications International Corp. status = planned datestring = Monday date = 2007-06-13

4.2.2.57   JointVenture

em/r/JointVenture	
Definition	A joint-venture agreement, signed recently or in the past.
Attributes	**company**: The company that is the subject of the relation type.
	company_newname: The new name of the company.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	status: The status, based on an indication in the text.

Example 1

Input Text	L-3 Communications (NYSE:LLL) and Thales Avionics, a wholly-owned subsidiary of Thales (formerly Thomson-CSF) specializing in avionics and aircraft cabin electronics, announced that they have signed an agreement to form a joint venture company to operate the assets of L-3's Aviation Communications & Surveillance Systems (ACSS).
Extracted Relation	company = L-3 Communications company = Thales Avionics status = announced

Example 2

Input Text	Building on the success of the MSNBC joint venture created by Microsoft and NBC in 1995, this new agreement brings together CNBC's powerful financial industry brand and programming expertise with Microsoft's extensive online distribution and technology experience.
Extracted Relation	company = Microsoft company = NBC status = known datestring = in 1995 date = 1995-00-00

4.2.2.58   ManMadeDisaster

em/r/ManMadeDisaster	
Definition	A disaster or incident that is not due to an act of nature. For example fire, nuclear disaster, oil spill, explosion.
Attributes	**date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	facility: The facility in which the incident occurred.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	location: Location related to the extracted relation (address or other physical description).
	manmadedisaster: Phrase describing a man-made disaster.

Example

Input Text	Rehabilitated Penguins are Released Following Oil Spill in Uruguay.
Extracted Relation	manmadedisaster = oil spill
	location = Uruguay

4.2.2.59   Merger

em/r/Merger	
Definition	A reference to a merger transaction (past, present, or future).
Attributes	**company**: Company involved in the merger.
	Note: The company attribute appears two or more times in the em/r/merger tag. Each company attribute indicates a different company that is involved in the merger.
	confidencelevel: A confidence score on a scale of 0 to 1. The value represents the probability that the extracted relation is indeed of the assigned type.
	For example:
	• The confidence level score for an em/r/deal tag indicates the probability that the deal is indeed a deal between the companies specified by the Acquirer and Target attributes.
	• The confidence level score for an em/r/bankruptcy tag indicates the probability that text actually refers to a bankruptcy involving the company specified by the Company attribute.
	• The confidence level score for an em/r/IPO tag indicates the probability that the text actually refers to an IPO involving the company specified by the Company attribute.
	The higher the value, the higher the probability.
	The consuming application can use this score to achieve higher accuracy results by ignoring instances with confidence scores below a specified level. Note that boosting Precision in this manner is at the expense of Recall.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	status: The status, based on an indication in the text.

Example

Input Text	First Union Corp., the nation's sixth-largest bank, said it will merge with Wachovia Corp.
Extracted Relation	company = First Union Corp. company = Wachovia Corp. status = planned

4.2.2.60   MilitaryAction

em/r/MilitaryAction	
Definition	A non-combative action taken by the military.
Attributes	**action:** Non-combative action taken by the military.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	locationstring: Location of the military action as mentioned in the text.
	military: A military group, or the country or person a military group serves.
	Note: When the name of a capital city is used to refer to a government, the attribute value will be the country name. In the following example, "Moscow deployed 10,000 extra troops," Russia is extracted as the Military attribute, and not Moscow.
	purpose: Purpose of the action.

Example 1

Input Text	Ministry head Shota Utiashvili said Moscow was deploying 10,000 extra troops in Georgia.
Extracted Relation	military = Russia action = was deploying 10,000 extra troops locationstring = Georgia

Example 2

Input Text	Thousands of Turkish troops have crossed into northern Iraq to hunt Kurdish rebels.
Extracted Relation	military = Thousands of Turkish troops action = have crossed into northern Iraq locationstring = Iraq

4.2.2.61   MovieRelease

em/r/MovieRelease	
Definition	A movie release (past, present, or future).
Attributes	**date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	movie: The movie.
	status: The status of the movie release based on an indication in the text.
	Possible values:
	• New (for a movie that was released during the current month or for a future release) • Known (all other instances, where there is no clear indication that it is a new or future release)

Example

Input Text	The latest Will Smith movie, "Hancock", will hit theaters Wednesday, July 2.
Extracted Relation	movie = Hancock datestring = Wednesday, July 2 date = 2008-07-02 status = New

4.2.2.62   MusicAlbumRelease

em/r/MusicAlbumRelease	
Definition	A music album release (past, present, or future).
Attributes	**date**: An absolute date (in the YYYY-MM-DD format). This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date or a reference to a date (e.g. yesterday, next month), taken directly from the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **musicalbum**: The album. **musicgroup_performer**: The musician or group releasing the album. **person_performer**: The person who is releasing the music album. **status:** The status of the filing based on an indication in the text. Possible values: • New (for an album that was released during the current month or for a future release) • Known (for all other instances, where there is no clear indication that it is a new or future release)

Example

Input Text	Country singer Jennifer Nettles, half of the duo Sugarland, entered the chart at No. 5 with her first solo album, "That Girl," selling 54,000 copies.
Extracted Relation	musicalbum = That Girl musicgroup_performer = Jennifer Nettles status = new

4.2.2.63   NaturalDisaster

em/r/NaturalDisaster	
Definition	A natural disaster. For example, a hurricane, an earthquake, a flood.
Attributes	**date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **location**: Location related to the extracted relation (address or other physical description). **naturaldisaster**: Phrase describing a natural disaster.

Example

Input Text	More than 250 people died when torrential rains and flooding swept China last month, according to the China Daily.
Extracted Relation	naturaldisaster = flood location = China date = 2009-02-00 datestring = last month

4.2.2.64   PatentFiling

em/r/PatentFiling	
Definition	A company or organization filing for patent registration, and information about the patent.
Attributes	**company**: The company that is the subject of the relation type.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	organization: The organization that is a participant in the extracted relation.
	patentdescription: Phrase describing the patent.
	status: The status of the filing based on an indication in the text.
	Possible values:
	• New (filed within the past 12 months)
	• Known (filed more than a year ago)

Example	
Input Text	Lixte Biotechnology Holdings announces filing of a New patent Application for Cancer Therapy Based on Targeting a Biomarker by Novel Investigational Agents Tuesday August 19, 8:15 am ET.
Extracted Relation	company = Lixte Biotechnology Holdings
	patentdescription = Cancer Therapy Based on Targeting a Biomarker by Novel Investigational Agents
	date = 2008-8-19
	datestring = Tuesday August 19
	status = New

4.2.2.65   PatentIssuance

em/r/PatentIssuance	
Definition	Reference to a patent being issued/granted to a company or organization, and information about the patent.
Attributes	**company**: The company that is the subject of the relation type.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	organization: The organization that is a participant in the extracted relation.
	patentdescription: Phrase describing the patent.
	patentnumber: The patent number.
	status: The status of the patent issuance based on an indication in the text.
	Possible values:
	• New (issued within the past 12 months)
	• Known (issued more than a year ago

Example

Input Text	Indevus Pharmaceuticals, Inc. announced today the issuance of U.S. Patent No. 7,410,978 (the "978 patent"), entitled "Once Daily Dosage Forms of Trospium," which is licensed exclusively to the Company by the patent's assignee, Supernus Pharmaceuticals, Inc.
Extracted Relation	company = Indevus Pharmaceuticals, Inc.
	patentnumber = 7,410,978 (the "978 patent")
	patentdescription = Once Daily Dosage Forms of Trospium
	date = 2005-07-15
	datestring = today
	status = New

4.2.2.66   PersonAttributes

em/r/PersonAttributes	
Definition	Extracts references to a person's age, birth date, birth place, and/or gender.
Attributes	**age**: The person's age. **birthdate**: The person's date of birth. **birthplace**: The person's place of birth. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **gender**: The person's gender. Possible values: "F" (female), "M" (male). **person**: The person who is the subject of this extracted relation.

Example 1

Input Text	Pennario was born in Buffalo on July 9, 1924. When he was 10 he and his family moved to Los Angeles
Extracted Relation	person = Leonard Pennario birthdate = 1924-07-09 birthplace = Buffalo gender = M

Example 2

Input Text	Mercedes-Benz Korea announced on Wednesday that Harald Behrend, 46, has been named as the new head of the company.
Extracted Relation	person = Harald Behrend age = 46

4.2.2.67   PersonCareer

em/r/PersonCareer	
Definition	A reference to a person and the position they hold or held in the past. The position may be that of a judge or other civil servant, a political office, or a professional position in a company or organization.
Attributes	**careertype**: Indicates whether the career is political or professional.
	city: City related to the extracted relation.
	company: Company affiliated with the position.
	country: Country related to the extracted relation.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	organization: The organization that is affiliated with the position.
	person: The person who is the subject of this extracted relation.
	position: Position that the person held, holds, or will hold.
	positionnormalized: A role in the legal profession.
	Possible values:
	• Attorney
	• Judge
	provinceorstate: The ProvinceOrState, related to the extracted relation.
	status: The status, based on an indication in the text.

Example

Input Text	DeLay, a Texas Republican with close ties to Taiwan's President Chen Shui-bian, went on the offensive on behalf of the investment bank.
Extracted Relation	person = Chen Shui-bian position = President country = Republic of China careertype = political status = current

4.2.2.68   PersonCommunication

em/r/PersonCommunication	
Definition	Communication (meeting, phone call, direct verbal communication, etc.) between people or groups of people. The meetings can be confirmed to have taken place, alleged to have taken place, or planned for the future. The meeting participants may be identified in this tag by any of the following attributes: person, persondescription, organizationorcompany, facility.
Attributes	**date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **facility**: The facility related to the extracted relation. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **organizationorcompany**: An organization or company participating in the communication. **person**: The person who is a participant in this extracted relation. **persondescription**: Description of participant/s in the communication. **status:** The status, based on an indication in the text.

Example

Input Text	Kerry will also meet Jordanian foreign minister Nasser Judeh in Rome.
Extracted Relation	person = John Kerry person = Judeh Nasser status = planned

4.2.2.69   PersonEducation

em/r/PersonEducation	
Definition	A person's academic degree/s or certification/s.
Attributes	**certification**: A mention of professional certification. **degree**: Academic degree awarded, not including the specific department or field of study. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **person**: The person who is the subject of this extracted relation. **schoolororganization**: Academic institution (typically a university or similar institution) that the person attended or attends, or that awarded the degree.

Example

Input Text	Mr. Liu graduated summa cum laude from UCLA with a B.A. in Economics in 1974.
Extracted Relation	person = Ron Liu degree = Bachelor of Arts schoolororganization = UCLA

4.2.2.70   PersonEmailAddress

em/r/PersonEmailAddress	
Definition	A person's e-mail address.
Attributes	**emailaddress:** The e-mail address. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **person**: The person who is the subject of this extracted relation.

Example

Input Text	**Cynthia Jackson**, +1-254-710-7628 (office), +1-254-749-4055 (cell), Cynthia_Jackson@baylor.edu
Extracted Relation	person = Cynthia Jackson emailaddress = Cynthia_Jackson@baylor.edu

4.2.2.71   PersonLocation

em/r/PersonLocation	
Definition	A person's location. The person may be visiting the location, or residing there permanently or temporarily.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **locationstring**: Location as mentioned in the text. **person**: The person who is the subject of this extracted relation. **persongroup**: A phrase that describes the participants in the extracted relation. For example, "twenty five Americans," "thousands of soldiers," "officials from the Ministry of Transportation," "the ambassador to Pakistan."

Example

Input Text	President Hamid Karzai met with reporters in Kabul.
Extracted Relation	persongroup = Hamid Karzai location = Kabul
Extracted Relation	person = reporters location = Kabul

4.2.2.72   PersonParty

em/r/PersonParty	
Definition	The political party affiliation of a person.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	party: The political party with which the person is associated. May be a full party name or a generic name such as Liberals, or Democrats.
	partydescription: A phrase from the text that describes the political party.
	person: The person belonging to or associated with a political party.

Example 1

Input Text	Republican candidates, including Rick Santorum and Newt Gingrich, attend a 2012 Primary Gala event, just days before votes are cast in the influential New Hampshire primary.
Extracted Relation	person = Rick Santorum party = Republican
Extracted Relation	person = Newt Gingrich party = Republican

Example 2

Input Text	French President Nicolas Sarkozy clashes with Socialist former prime minister Laurent Fabius in a first face-to-face debate.
Extracted Relation	person = Laurent Fabius party = Socialist

Example 3

Input Text	Communist Party of Nepal (Maoist) -- founded in 1994 by Pushpa Kamal Dahal, widely known as Prachanda.
Extracted Relation	person = Pushpa Kamal Dahal party = Communist Party of Nepal partydescription = Maoist

4.2.2.73   PersonRelation

em/r/PersonRelation	
Definition	A business, friend, academic, military-service, political, or romantic relationship between people.
Attributes	**date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **person**: The person who is a participant in this extracted relation. **personalrelationtype**: Classifies the relationship between two people. Possible values: • Business • Friends • Academic • Military_Service • Politics • Romantic.

Example

Input Text	Mark McDonald and Ray Odierno served together in combat in 2008, and McDonald said Odierno was an "absolute joy" to work with.
Extracted Relation	person = Mark McDonald person = Ray Odierno datestring = in 2008 date = 2008-00-00 personrelationtype = Military_service

4.2.2.74   PersonTravel

em/r/PersonTravel	
Definition	A mention of a person travelling, past, present, or future.
Attributes	**date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	locationdestination: The travel destination.
	locationorigin: The place of travel origin.
	person: The person who is the subject of this extracted relation.
	status: The status, based on an indication in the text.

Example

Input Text	In another sign of easing tensions, Indian officials said that Prime Minister Atal Bihari Vajpayee might visit Pakistan early next year for a regional summit.
Extracted Relation	person = Atal Bihari Vajpayee
	locationdestination = Pakistan
	datestring = next year
	date = 2003-00-00
	status = possible_future

4.2.2.75   PoliticalEndorsement

em/r/PoliticalEndorsement	
Definition	An endorsement of a political entity (person, party, lobby) by another political entity.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **groupendorsee**: The political entity (i.e. person, party, lobby) that received the endorsement. **groupendorser**: The political entity (i.e. person, party, lobby) that made the endorsement.

Example

Input Text	Obama wins endorsement of government employees union.
Extracted Relation	groupendorsee = Barack Obama groupendorser = government employees union

4.2.2.76   PoliticalRelationship

em/r/PoliticalRelationship	
Definition	A relationship (alliance or rivalry) between two political entities (people, political parties, lobbies).
	Note: Endorsements are not extracted as PoliticalRelationships.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **politicalentity1**: First political entity of the relationship. **politicalentity2**: Second political entity in the relationship. **politicalrelationshiptype**: Classifies the relationship. Possible values: • Alliance • Rivalry

Example

Input Text	The Democrats' expected coalition partners include the Liberal Democratic Party, whose leader Cedomir Jovanovic continued to his second term.
Extracted Relation	politicalentity1 = Democrats politicalentity2 = Liberal Democratic Party politicalrelationshiptype = Alliance

4.2.2.77   PollsResult

em/r/PollsResult	
Definition	Indication of voting results as determined by a poll.
Attributes	**date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	location: Location related to the extracted relation (address or other physical description).
	opponent: The opposing organization or party.
	politicalevent: The political event.
	pollconductor: Name of the company/organization conducting the poll.
	winningcandidate: The name of the winning person or political party.

Example

Input Text	(AP) Exit polls showed conservative challenger Angela Merkel's party leading in German parliamentary elections Sunday.
Extracted Relation	politicalevent = parliamentary elections winningcandidate = Angela Merkel pollconductor = AP date = 2005-09-18 datestring = Sunday

4.2.2.78   ProductIssues

em/r/ProductIssues	
Definition	An issue or problem with a product. This tag may be generated if product issues are mentioned in association with a company, even if no specific product is mentioned.
Attributes	**company**: The company that is the subject of the relation type. **date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **product**: The product that is the subject of the extracted relation. **producttype**: Product classification. Possible values: • aircraft • car • drug • electronics • weapon • other **status:** The status, based on an indication in the text.

Example 1

Input Text	Consumers are being warned to stop using the blinds immediately.
Extracted Relation	product = blinds producttype = Other status = known

Example 2

Input Text	Actavis Totowas LLC., manufacturer of the recalled defective heart drug Digitek, has announced a voluntary recall of all drug products manufactured in its Little Falls, NJ facility.
Extracted Relation	product = Digitek producttype = Drug company = Actavis Totowas LLC status = announced

4.2.2.79 ProductRecall

em/r/ProductRecall	
Definition	A product recall.
Attributes	**company_distributor**: Company that distributes the product, or the name of the manufacturer's agent.
	company_recalling: Company that is recalling a product.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	product: The product that is the subject of the extracted relation.
	producttype: Product classification.
	Possible values:
	• aircraft
	• car
	• drug
	• electronics
	• weapon
	• other
	recalledquantity: The recalled quantity.
	status: The status, based on an indication in the text.

Example 1

Input Text	Toyota is recalling about 533,000 of its Sequoia sports cars.
Extracted Relation	company_recalling = Toyota Motors Corporation
	product = Sequoia
	producttype = Car
	recalledquantity = about 533,000
	status = announced

Example 2

Input Text	Supervalu Inc. (NYSE:SVU) announced yesterday the precautionary recall of trail mix and fruit and nut trail mix products sold under the Acme, Albertsons, Jewel and Shaw brands.
Extracted Relation	company_recalling = Supervalu Inc.
	product = trail mix
	producttype = Other
	date = 2009-03-08
	datestring = yesterday

4.2.2.80   ProductRelease

em/r/ProductRelease	
Definition	A reference to a product release.
Attributes	**company**: The company that is the subject of the relation type.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	product: The product that is the subject of the extracted relation.
	producttype: Product classification.
	Possible values:
	• aircraft
	• car
	• drug
	• electronics
	• weapon
	• other
	status: The status, based on an indication in the text.

Example

Input Text	Microsoft Corp. said Thursday it will delay the release of its Office 2007 Applications Suite.
Extracted Relation	company = Microsoft Corporation product = Office 2007 Applications Suite producttype = Other date = 2006-06-27 datestring = Thursday status = delayed

4.2.2.81   Quotation

em/r/Quotation	
Definition	A direct or indirect quotation.
Attributes	**date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **quotation**: Direct or indirect quotation, taken directly from the text. **quotationtype**: Classifies the quotation as direct or indirect. Possible values: • Primary (the mention appears within quotation marks) • Paraphrase (the mention does not appear within quotation marks) **speaker**: The person, company, organization, or country being quoted.

Example

Input Text	"It's going to have a real impact on the level of services," Acting Treasurer David Rousseau said at a briefing for reporters.
Extracted Relation	speaker = David Rousseau quotation = It's going to have a real impact on the level of services quotationtype = Primary

4.2.2.82   SecondaryIssuance

em/r/SecondaryIssuance	
Definition	Stock issuance by a company that is not an initial issuance. (Initial issuance of stock generates the em/r/IPO tag.)
Attributes	**company**: The company that is the subject of the relation type.
	date: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	status: The status, based on an indication in the text.

Example

Input Text	Lumera Corporation plans to raise $17 mln from private offer.
Extracted Relation	company = Lumera Corporation status = planned

4.2.2.83   StockSplit

em/r/StockSplit	
Definition	A stock split declared by a company.
Attributes	**company**: The company that is the subject of the relation type.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	stocksplitratio: The ratio of the number of shares before the split to the number of shares after the split, if indicated in the text.

Example 1

Input Text	The Board of Directors of Curtiss-Wright Corporation (NYSE: CW; CW.B) today declared a two-for-one split of the company's Common Stock.
Extracted Relation	company = Curtiss-Wright Corporation stocksplitratio = 2:1

Example 2

Input Text	Poniard Pharmaceuticals to Seek Shareholder Approval for Reverse Stock Split
Extracted Relation	company = Poniard Pharmaceuticals

4.2.2.84   Trial

em/r/Trial	
Definition	A trial of one or more people.
Attributes	**charge**: Phrase describing an offense/charge. In case of multiple charge attributes, the first charge attribute is the primary charge. **date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date. **datestring**: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **othercharges**: Additional charges. **person**: The person who is the subject of this extracted relation.

Example 1

Input Text	Judge Brinkema announced the schedule for the penalty phase of Zacarias Moussaoui's trial Friday.
Extracted Relation	person = Zacarias Moussaoui datestring = Friday date = 2009-03-06

Example 2

Input Text	Jermaine Wallace, 25, of Salisbury, and Natasha Adams, 25, of Bowie, were charged with conspiracy to distribute cocaine and money laundering.
Extracted Relation	person = Jermaine Wallace charge = conspiracy to distribute cocaine charge = money laundering
Extracted Relation	person = Natasha Adams charge = conspiracy to distribute cocaine charge = money laundering

4.2.2.85 VotingResult

em/r/VotingResult	
Definition	Information about the results of a vote.
Attributes	**date**: An absolute date (in the YYYY-MM-DD format) that is related to the extracted relation. This date is either mentioned in the text, or calculated based on the datestring attribute value and relative to the text publication date.
	datestring: A mention of a date, or a reference to a date (e.g. yesterday, next month) relating to the extracted relation, as it appears in the text.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	location: Location related to the extracted relation (address or other physical description).
	opponent: The opposing organization or party.
	politicalevent: The political event.
	winningcandidate: The name of the winning person or political party.

Example

Input Text	Boris Johnson defeated Ken Livingstone in the London mayoral election.
Extracted Relation	politicalevent = mayoral election winningcandidate = Boris Johnson opponent = Ken Livingstone location = London

4.3 RelevanceInfo Tag

For a conceptual explanation of this tag, see Relevance Tags.

RelevanceInfo	
Definition	Indicates how centric the associated entity is to the document.
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	subject: A hash tag generated by Open Calais. This ID points to the associated extracted entity.
	Note: In the JSON output format, because related tags are nested, the docid and subject attributes are not displayed in the RelevanceInfo tag.
	relevance: The relevance score. The higher the score, the greater the relevance of the entity to the document.
	Currently, the following values are supported:
	• 1.0 – This value is reserved for the company identified as the reporting company in a document with a predefined format such as an SEC report.
	• 0.8 – Entities defined as having high relevance will receive this score.
	• 0.5 – Entities defined as having medium relevance will receive this score.
	• 0.2 – Entities defined as having low relevance will receive this relevance score.
	• 0.0 – This value is reserved for entities identified as irrelevant to the story. For example, mentions of companies as rating agencies, reporting agencies, stock exchanges, and social applications receive a relevance score of 0.
	relevancecont: A more granular relevance score. This attribute is relevant only to Company entities.
	This attribute is not relevant to JSON output.
	• A score that is 0.8 and above indicates a high relevance.
	• A score of 0 indicates zero relevance.
	For both **relevance** and **relevancecont** scores, the three following buckets are the most indicative ones:
	• High relevance: score is equal to or greater than 0.8
	• Zero relevance: score equals 0
	• Everything else: score is greater than zero and less than 0.8
	Note: For the highest accuracy, we recommend ranking companies based on the **relevance** score and *not* on the **relevancecont** score

RelevanceInfo

Example

Open Calais Output	*RDF...* ```<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6/Relevance/47">` ` <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/RelevanceInfo"/>` ` <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>` ` <c:subject rdf:resource="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4"/>` ` <c:relevance>0.8</c:relevance>` ` <c:relevancecont>0.72</c:relevance>` `</rdf:Description>``` *N3...* ```<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Relevance/47>` ` a <http://s.opencalais.com/1/type/sys/RelevanceInfo> ;` ` c:docId <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;` ` c:relevance "0.8" ;` ` c:relevancecont "0.72" ;` ` c:subject <http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4> .``` *JSON...* ``` "relevance": 0.8` `},```

4.4 Confidence Tag

Currently, Open Calais generates Confidence tags for the following entity and relation types: Company, Person, Pharmaceutical Drug, Bankruptcy, Deal, IPO.

The consuming application can achieve higher accuracy results by ignoring instances with confidence scores below a specified level. Note, however, that raising the specified level boosts Precision at the expense of Recall.

For a conceptual explanation of this tag, see Confidence Tags.

Confidence	
Definition	Indicates the probability that the associated e.g. extracted person or company is indeed a person or company.
Attributes	**aggregate**: The confidence score. **dblookup**: A value used internally by Open Calais to calculate the confidence score. **docid**: The unique ID of the containing document, generated by Open Calais, and formatted as a URI. **resolution**: A value used internally by Open Calais to calculate the confidence score. **statisticalfeature**: A value used internally by Open Calais to calculate the confidence score. **subject**: A hash tag generated by Open Calais. This ID points to the associated extracted entity or relation. **Note:** In the JSON output format, because related tags are nested, the docid and subject attributes are not displayed in the Confidence tag.

Confidence

Example

Open Calais Output	*RDF...*
	``<rdf:Description rdf:about="http://d.opencalais.com/conf/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4">
<rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/Confidence"/>
<c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>
<!--Apple-->
<c:subject rdf:resource="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4"/>
<c:statisticalfeature>0.997</c:statisticalfeature>
<c:dblookup>0.0</c:dblookup>
<c:resolution>0.9928677</c:resolution>
<c:aggregate>0.996</c:aggregate>
</rdf:Description>```
	N3...
	``<http://d.opencalais.com/conf/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4>
a   <http://s.opencalais.com/1/type/tag/Confidence> ;
c:aggregate   "0.996" ;
c:dblookup   "0.0" ;
c:docId   <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;
c:resolution   "0.9928677" ;
c:statisticalfeature   "0.997" ;
c:subject   <http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4> .```
	JSON...
	``"confidence": {
"statisticalfeature": "0.997",
"dblookup": "0.0",
"resolution": "0.9928677",
"aggregate": "0.996"
},```

4.5   Disambiguation Tags

Currently, Open Calais supports Disambiguation for the following entity and relation types:

[City](#)

[Company](#)

[Continent](#)

[Country](#)

[Currency](#)

[Deal](#)

[Organization](#)

[Person](#)

[Product](#)

[ProvinceOrState](#)

[Region](#)

[TopmostPublicParentCompany](#)

Click a link for a description of the tag and its attributes.

Note: It is important to note that most attributes are optional; a tag can be extracted with some but not all of its attributes.

For a conceptual explanation of this tag, see [Disambiguation Tags](#).

4.5.1 City

er/Geo/City	
Definition	This tag is generated when an extracted entity of the type City is successfully mapped to a city in the Thomson Reuters Classification Schema (TRCS).
Attributes	**containedbycountry**: The name of the country in which the geographical entity is located.
	containedbystate: The name of the state in which the geographical entity is located.
	docid: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	latitude: The latitude of the geographical entity, as defined in Freebase.
	longitude: The longitude of the geographical entity, as defined in Freebase.
	name: The official name as it appears in the Thomson Reuters dataset.
	permid: The Thomson Reuters unique ID (PermID) for this geographical entity.
	rcscode: The code of this geographical entity, as defined in the Thomson Reuters Coding Schema (TRCS).
	shortname: A name commonly used to refer to the geographical entity, as defined in Freebase.
	subject: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/City

4.5.2   Company

er/Company	
Definition	This tag is generated when an extracted entity of the type Company is mapped to a company in the Thomson Reuters dataset. Note that the *openpermid* attribute gives you direct access to high quality, curated Thomson Reuters company data. (In JSON output, the attribute name is *id*.)
Attributes	**commonname**: The commonly used name, as it appears in the Thomson Reuters dataset. **docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **id:** See the *openpermid* attribute description, below. **ispublic**: Indicates whether the company is public (true) or private (false). **legacyid**: The link to the company as it would have been displayed in the legacy Open Calais er/company tag. This attribute is relevant to users who upgraded from the legacy Open Calais to the new Open Calais. The purpose of this attribute is to enable you to map the old links, which are no longer supported, to the corresponding new links. This attribute is not relevant to the JSON output format. An example of an old company link in the legacy Open Calais er/company tag (RDF output): ``` <rdf:Description rdf:about="http://d.opencalais.com/er/company/ralg-tr1r/39d9a86f-6241-3688-b05b-4595eeaba2d1"> ``` The company link and the legacyid attribute as they appear in the new Open Calais er/company tag (RDF output): ``` <rdf:Description rdf:about="http://d.opencalais.com/er/company/ralg-oa/5000608903"> <c:legacyid rdf:resource="http://d.opencalais.com/er/company/ralg-tr1r/39d9a86f-6241-3688-b05b-4595eeaba2d1"/> ``` The company link and the legacyid attribute as they appear in the new Open Calais er/company tag (N3 output): ``` <http://d.opencalais.com/er/company/ralg-oa/5000608903> c:legacyid < http://d.opencalais.com/er/company/ralg-tr1r/39d9a86f-6241-3688-b05b-4595eeaba2d1> ; ``` **name**: The official name as it appears in the Thomson Reuters dataset. **openpermid:** A direct link to the relevant company page on the Open PermID website (https://permid.org). This link gives you direct access to high quality, curated Thomson Reuters company data. (In the JSON output, the attribute name is **id**.) **permid:** The Thomson Reuters unique ID (PermID) for this company. **primaryric**: The company's primary Reuters Instrument code (RIC) as it appears in the Thomson Reuters dataset. **score**: A score representing the probability that the extracted company was accurately mapped to the corresponding company in the Thomson Reuters dataset. Scores range from 0 - 1; the higher the score, the higher the probability. The consuming application can use this score to achieve higher accuracy results. **subject**: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format. **ticker**: The company stock symbol.

Related Tags	em/e/Company
	er/TopmostPublicParentCompany

4.5.3   Continent

er/Continent	
Definition	This tag is generated when an extracted entity of the type Continent is successfully mapped to a continent in the Thomson Reuters Classification Schema (TRCS).
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **name**: The official name as it appears in the Thomson Reuters dataset. **permid**: The Thomson Reuters unique ID (PermID) for this geographical entity. **rcscode**: The code of this geographical entity, as defined in the Thomson Reuters Coding Schema (TRCS). **subject**: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Continent

4.5.4   Country

er/Geo/Country	
Definition	This tag is generated when an extracted entity of the type Country is successfully mapped to a country in the Thomson Reuters Classification Schema (TRCS).
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **latitude**: The latitude of the geographical entity, as defined in Freebase. **longitude**: The longitude of the geographical entity, as defined in Freebase. **name**: The official name as it appears in the Thomson Reuters dataset. **permid**: The Thomson Reuters unique ID (PermID) for this geographical entity. **rcscode**: The code of this geographical entity, as defined in the Thomson Reuters Coding Schema (TRCS). **shortname**: A name commonly used to refer to the geographical entity, as defined in Freebase. **subject**: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Country

4.5.5   Currency

er/Currency	
Definition	This tag is generated when an extracted entity of the type Currency is successfully mapped to a currency in the Thomson Reuters Classification Schema (TRCS).
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **name**: The official name as it appears in the Thomson Reuters dataset. **permid**: The Thomson Reuters unique ID (PermID) for this currency. **rcscode**: The currency code, as defined in the Thomson Reuters Coding Schema (TRCS). **subject**: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Currency

4.5.6   Deal

This functionality is available upon subscription to Thomson Reuters Intelligent Tagging (TRIT). For more information please contact us at questions@opencalais.com.

er/Deal	
Definition	This tag is generated when an extracted relation of the type Deal is successfully mapped to a deal in the Thomson Reuters Deals Authority dataset.
Attributes	**dealsdcnum**: The deal ID as defined in the Thomson Reuters Deals Authority dataset. **docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **permid**: The Thomson Reuters unique ID (PermID) for this deal. **subject**: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/r/Deal

4.5.7 Organization

er/Organization	
Definition	This tag is generated when an extracted entity of the type Organization is successfully mapped to an organization in the Thomson Reuters Organization Authority dataset.
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	name: The official name as it appears in the Thomson Reuters dataset.
	permid: The Thomson Reuters unique ID (PermID) for this organization.
	rcscode: The organization code, as defined in the Thomson Reuters Coding Schema (TRCS).
	subject: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Organization

4.5.8   Person

er/Person	
Definition	This tag is generated when an extracted entity of the type Person is successfully mapped to a person in the Thomson Reuters People Authority dataset. Currently this tag is only generated for executives.
Attributes	**commonname**: The commonly used name, as it appears in the Thomson Reuters dataset.
	docid: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	name: The official name as it appears in the Thomson Reuters dataset.
	officerid: A unique ID (as defined in the Thomson Reuters Officers and Directors dataset) that specifies a position held (currently or in the past) by the person, in a particular company. The er/person tag may include multiple officerid attributes.
	openpermid: A direct link to the relevant person page on the Open PermID website (https://permid.org). This link gives you direct access to high quality, curated Thomson Reuters people data. (In the JSON output, the attribute name is **id**.)
	paid: The Thomson Reuters unique ID (from the People Authority) for this person.
	personid: The Person ID as defined in the Thomson Reuters Officers and Directors dataset.
	score: A score representing the probability that the extracted person was accurately mapped to the corresponding person in the Thomson Reuters dataset.
	Scores range from 0 - 1; the higher the score, the higher the probability. The consuming application can use this score to achieve higher accuracy results.
	subject: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Person

4.5.9   Product

er/Product	
Definition	This tag is generated when an extracted entity of the type Product is successfully mapped to a product in the Thomson Reuters dataset.
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **name**: The official name as it appears in the Thomson Reuters dataset. **ownername**: The name of the organization that owns the product. The organization name is as it appears in the Thomson Reuters dataset. **ownerpermid:** The permid of the organization that owns the product. **subject**: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Product

4.5.10   ProvinceOrState

er/Geo/ProvinceOrState	
Definition	This tag is generated when an extracted entity of the type ProvinceOrState is successfully mapped to a province or state in the Thomson Reuters Classification Schema (TRCS).
Attributes	**containedbycountry**: The name of the country in which the geographical entity is located.
	containedbystate: The name of the state in which the geographical entity is located.
	docid: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	latitude: The latitude of the geographical entity, as defined in Freebase.
	longitude: The longitude of the geographical entity, as defined in Freebase.
	name: The official name as it appears in the Thomson Reuters dataset.
	permid: The Thomson Reuters unique ID (PermID) for this geographical entity.
	rcscode: The code of this geographical entity, as defined in the Thomson Reuters Coding Schema (TRCS).
	shortname: A name commonly used to refer to the geographical entity, as defined in Freebase.
	subject: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/ProvinceOrState

4.5.11   Region

er/Region	
Definition	This tag is generated when an extracted entity of the type Region is successfully mapped to a region in the Thomson Reuters Classification Schema (TRCS).
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **name**: The official name as it appears in the Thomson Reuters dataset. **permid**: The Thomson Reuters unique ID (PermID) for this region. **rcscode**: The code of this region, as defined in the Thomson Reuters Coding Schema (TRCS). **subject**: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Region

4.5.12 TopmostPublicParentCompany

er/TopmostPublicParentCompany	
Definition	This tag is generated when an extracted entity of the type Company is mapped to a company in the Thomson Reuters dataset that also has a publicly traded parent company.
	Note that the *openpermid* attribute gives you direct access to high quality, curated Thomson Reuters company data. (In JSON output, the attribute name is *id*.)
	This metadata type is available upon subscription to Thomson Reuters Intelligent Tagging (TRIT). For more information please contact us at questions@opencalais.com.
Attributes	**commonname**: The commonly used name, as it appears in the Thomson Reuters dataset.
	docid: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	id: See the *openpermid* attribute description, below.
	ispublic: Indicates whether the company is public (true) or private (false).
	legacyid: The link to the company as it would have been displayed in the legacy Open Calais output.
	This attribute is relevant to users who upgraded from the legacy Open Calais to the new Open Calais.
	The purpose of this attribute is to enable you to map the old links, which are no longer supported, to the corresponding new links.
	This attribute is not relevant to the JSON output format.
	name: The official name as it appears in the Thomson Reuters dataset.
	openpermid: A direct link to the relevant company page on the Open PermID website (https://permid.org). This link gives you direct access to high quality, curated Thomson Reuters company data. (In the JSON output, the attribute name is **id**.).)
	permid: The Thomson Reuters unique ID (PermID) for this company.
	primaryric: The company's primary Reuters Instrument code (RIC) as it appears in the Thomson Reuters dataset.
	subject: The hash tag generated by Open Calais and assigned to the extracted child company. This ID points to the relevant em/e/company and related tags within the document. If the parent company has multiple extracted child companies within the document, additional instances of the subject attribute appear in the tag. Note that this entity is not relevant to the JSON output format.
	ticker: The company stock symbol.
Related Tag	em/e/Company

4.6   SocialTag

For a conceptual explanation, see Social Tags.

SocialTag	
Definition	Classifies the document based on Wikipedia folksonomy.
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **importance**: Indicates how centric the topic identified by the social tag is to the document as a whole. Possible values: **1** (very centric); **2** (somewhat centric); **3** (less centric). **name**: A word or phrase that describes the document as a whole. (The title of the associated Wikipedia article.) **originalvalue**: The original title of the associated Wikipedia article, if the title has changed. **socialTag**: The unique identifier for the socialTag across documents.

SocialTag

Example

Open Calais Output	*RDF...* ```<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6/SocialTag/4">```

RDF...

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
    bf3520cddfa6/SocialTag/4">
    <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/SocialTag"/>
    <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
    bf3520cddfa6"/>
    <c:socialtag rdf:resource="http://d.opencalais.com/genericHasher-1/083d56d1-2fed-3063-
    b59a-963d4fdaee36"/>
    <c:forenduserdisplay>true</c:forenduserdisplay>
    <c:name>Electric car</c:name>
    <c:importance>2</c:importance>
    <c:originalValue>Electric car</c:originalValue>
</rdf:Description>
```

N3...

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/SocialTag/4>
        a                       <http://s.opencalais.com/1/type/tag/SocialTag> ;
        c:docId                 <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-
        816cb722a630> ;
        c:forenduserdisplay   "true" ;
        c:importance          "2" ;
        c:name                "Electric car" ;
        c:originalValue       "Electric car" ;
        c:socialtag           <http://d.opencalais.com/genericHasher-1/083d56d1-2fed-3063-b59a-
        963d4fdaee36> .
```

JSON...

```
"http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/SocialTag/4": {
        "_typeGroup": "socialTag",
        "id": "http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-
        816cb722a630/SocialTag/4",
        "socialTag": "http://d.opencalais.com/genericHasher-1/083d56d1-2fed-3063-b59a-
        963d4fdaee36",
        "forenduserdisplay": "true",
        "name": "Electric car",
        "importance": "2",
        "originalValue": "Electric car"
    },
```

4.7 Topic Tag (DocCat)

For a conceptual explanation, see Topic Tags.

DocCat	
Definition	Identifies the topics discussed in the document. The list of possible topics is defined by the Thomson Reuters Coding Schema (TRCS) and/or by the International Press Telecommunications Council (IPTC) news taxonomy. See List of Supported Topics. Access to TRCS topics is available upon subscription to Thomson Reuters Intelligent Tagging (TRIT). For more information please contact us at questions@opencalais.com.
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The topic name. This attribute is relevant to TRCS and Calais Self Service Classification topics only. **permid**: The Thomson Reuters unique ID (PermID) for this topic. **rcscode**: The topic code, as defined in the Thomson Reuters Coding Schema (TRCS). **score**: A confidence score on a scale of 0 to 1. The value indicates the probability that the topic is indeed discussed in the text and also how centric the topic is to the text. The higher the value, the higher the probability. The consuming application can use this score to achieve higher accuracy results by ignoring instances with scores below a specified level. **shortName**: Output for TRCS topics only. If the **name** attribute includes the string "**(TRBC)**," then the **shortName** attribute is the same, minus "**(TRBC)**." If the name attribute does not include the "(TRBC)" string, then the name and shortName attributes are identical. For example, if the topic **name** is "Technology Equipment (TRBC)," then the **shortName** is "Technology Equipment."

DocCat

Example

Open Calais Output	*RDF...*
	An IPTC taxonomy topic:

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
      bf3520cddfa6/cat/1">
      <rdf:type rdf:resource="http://s.opencalais.com/1/type/cat/DocCat"/>
      <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
      bf3520cddfa6"/>
      <c:forenduserdisplay>false</c:forenduserdisplay>
      <c:score>0.988</c:score>
      <c:name>Technology_Internet</c:name>
</rdf:Description>
```

A TRCS taxonomy topic:

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-
      0103f889be3e/cat/2">
      <rdf:type rdf:resource="http://s.opencalais.com/1/type/cat/DocCat"/>
      <c:docId rdf:resource="http://d.opencalais.com/dochash-1/f4707556-c36e-39af-b0e6-
      0103f889be3e"/>
      <c:forenduserdisplay>false</c:forenduserdisplay>
      <c:rcscode>B:162</c:rcscode>
      <c:name>Technology Equipment</c:name>
      <c:shortName>Technology Equipment</c:shortName>
      <c:score>0.273</c:score>
</rdf:Description>
```

N3...

An IPTC taxonomy topic:

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/cat/1>
      a                    <http://s.opencalais.com/1/type/cat/DocCat> ;
      c:docId              <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-
      816cb722a630> ;
      c:forenduserdisplay  "false" ;
      c:name               "Technology_Internet" ;
      c:score              "0.988" .
```

JSON...

An IPTC taxonomy topic:

```
"http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/cat/1": {
      "_typeGroup": "topics",
      "forenduserdisplay": "false",
      "score": 0.988,
      "name": "Technology_Internet"
},
```

4.8   Industry Tag

This functionality is available upon subscription to Thomson Reuters Intelligent Tagging (TRIT). For more information please contact us at questions@opencalais.com.

For a conceptual explanation, see Industry Tags.

Industry	
Definition	Identifies the industries related to the companies mentioned in the document.
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The name of the industry.
	permid: The Thomson Reuters unique ID (PermID) for this Industry.
	rcscode: The industry code, as defined in the Thomson Reuters Coding Schema (TRCS).
	relevance: Indicates how relevant the industry is to the story. Values range from 0 to 1. The higher the score, the higher the relevance.
	trbccode: The industry code, as defined in the Thomson Reuters Business Classification (TRBC) taxonomy.

Industry

Example

Open Calais Output	*RDF...*

```
    <rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
        bf3520cddfa6/Industry/5">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/Industry"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
        bf3520cddfa6"/>
        <c:forenduserdisplay>false</c:forenduserdisplay>
        <c:name>Automobiles &amp; Multi Utility Vehicles</c:name>
        <c:rcscode>B:1294</c:rcscode>
        <c:permid>4294951707</c:permid>
        <c:relevance>0.500</c:relevance>
    </rdf:Description>
```

N3...

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Industry/5>
        a                       <http://s.opencalais.com/1/type/tag/Industry> ;
        c:docId                 <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-
        816cb722a630> ;
        c:forenduserdisplay     "false" ;
        c:name                  "Automobiles & Multi Utility Vehicles" ;
        c:permid                "4294951707" ;
        c:rcscode               "B:1294" ;
        c:relevance             "0.500" .
```

JSON...

```
"http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Industry/5": {
        "_typeGroup": "industry",
        "forenduserdisplay": "false",
        "name": "Automobiles & Multi Utility Vehicles",
        "rcscode": "B:1294",
        "permid": "4294951707",
        "relevance": 0.5
    },
```

Chapter 5   French Semantic Metadata Tags

Open Calais is able to extract a rich set of metadata tags from French input text:

- InstanceInfo Tags

- The following Entity Markup (em/e) tags: City, Company, Continent, Country, Currency, EmailAddress, FaxNumber, MarketIndex, NaturalFeature, Organization, Person, PhoneNumber, ProvinceOrState, Region, URL

- RelevanceInfo Tags

- Confidence Tags

- The following Disambiguation (er) tags: City, Company, Continent, Country, Currency, Organization, Person, ProvinceOrState, Region, TopmostPublicParentCompany

- Industry Tags

(Currently, Social Tags and Topic Tags are not extracted from French input text.)

For a conceptual overview of the Open Calais Semantic Metadata Tags, see How Does Open Calais Work?

5.1 InstanceInfo Tag

For a conceptual explanation of this tag, see Instance Tags.

InstanceInfo	
Definition	Describes a mention of an Open Calais type found in the text.
Attributes	**detection**: The text string in which the mention was identified.
	docid: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	exact: The mention.
	length: The length (in characters) of the mention.
	offset: Offset of the mention (in characters) from the beginning of the document.
	prefix: The portion of the text string that precedes the mention.
	subject: A hash tag generated by Open Calais. This ID points to the associated extracted entity or relation.
	suffix: The portion of the text string that follows the mention.
	The tag also includes a comment that indicates the type of entity or relation extracted from this instance, and its main attribute values.
	Note: In the JSON output format, because related tags are nested, the docid and subject attributes are not displayed in the Instance tag.

Example:

Input Text	MONTREAL - La société Clarke (TSX:CKI), basée à Halifax, a présenté une offre pour racheter la compagnie montréalaise de vidéos et de musique Madacy Entertainment (TSX: MEG.UN) et en fermer le capital.

Instance Tag	<pre><rdf:Description rdf:about="http://d.opencalais.com/dochash-1/c70dbc7b-6035-38bb-baee-5644f634ae1d/Instance/1">
	   <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/InstanceInfo"/>
	   <c:docId rdf:resource="http://d.opencalais.com/dochash-1/c70dbc7b-6035-38bb-baee-5644f634ae1d"/>
	   <c:subject rdf:resource="http://d.opencalais.com/comphash-1/8724d616-faba-36c4-b140-8dfa315e8385"/>
	   <!--Company: Clarke; -->
	   <c:detection>[&lt;Body&gt; MONTREAL - La société]Clarke[(TSX:CKI), basée à Halifax, a présenté une offre]</c:detection>
	   <c:prefix>&lt;Body&gt; MONTREAL - La société </c:prefix>
	   <c:exact>Clarke</c:exact>
	   <c:suffix> (TSX:CKI), basée à Halifax, a présenté une offre</c:suffix>
	   <c:offset>95</c:offset>
	   <c:length>6</c:length>
	</rdf:Description></pre>

5.2   Entity Markup Tags

Click a link for a description of the entity type and its attributes.

City, Company, Continent, Country, Currency, EmailAddress, FaxNumber, MarketIndex, NaturalFeature, Organization, Person, PhoneNumber, ProvinceOrState, Region, URL

Note: It is important to note that most attributes are optional; a tag can be extracted with some but not all of its attributes.

For a conceptual explanation, see Entity Markup Tags.

5.2.1   City

em/e/City	
Definition	The name of a city or other municipality. Districts or neighborhoods within a city (e.g. Soho, London) are not extracted.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
Related Tag	er/Geo/City (Disambiguation Tag)

Example1

Input Text	Il y a 3 jours, elle était à Cannes où elle s'était produite au Palais Club.
Extracted Entity	Cannes

Example2

Input Text	Jacques Chirac, vêtu comme un estivant et détendu comme un touriste, visite Saint-Tropez.
Extracted Entity	Saint-Tropez

5.2.2   Company

em/e/Company	
Definition	A full or partial company name. The term "Company" refers to any business organization, including newspapers, media companies, law firms, etc.
Attributes	**confidencelevel**: A confidence score on a scale of 0 to 1. The value represents the probability that the extracted entity is indeed of the assigned type. For example: • The confidence level score for an em/e/person tag indicates the probability that the extracted person is indeed a person. • The confidence level score for an em/e/company tag indicates the probability that the extracted company is indeed a company. • The confidence level score for an em/e/pharmaceuticalDrug tag indicates the probability that the extracted pharmaceutical drug is indeed a pharmaceutical drug. The higher the value, the higher the probability. The consuming application can use this score to achieve higher accuracy results by ignoring instances with confidence scores below a specified level. Note that boosting Precision in this manner is at the expense of Recall. Every em/e tag that displays this attribute also has a corresponding Confidence tag. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). The related confidence score determines if the forenduserdisplay value is true or false. **inlineric**: The company Reuters Instrument Code (RIC), if it is mentioned in the text. **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used. **nationality**: The nationality of the extracted Person, Company, or Organization, if it is mentioned in the text.
Related Tags	Confidence, RelevanceInfo, er/Company (Disambiguation Tag)

Example1

Input Text	MONTREAL - La société Clarke (TSX:CKI), basée à Halifax, a présenté une offre pour racheter la compagnie montréalaise de vidéos et de musique Madacy Entertainment (TSX: MEG.UN) et en fermer le capital.
Extracted Entity	Clarke Madacy Entertainment

Example2

Input Text	A Xingtai, dans la province d'Hebei, la société Jinniu Energya incinéré quelque 1.200 tonnes de lait en poudre à plus de 1.800 degrés le mois dernier.
Extracted Entity	Jinniu Energya

5.2.3   Continent

em/e/Continent	
Definition	A continent.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The entity name, based on a mention in the text and mapped to a value from an internal reference list.
Related Tag	er/Continent (Disambiguation Tag)

Example

Input Text	L'Allemagne reste le premier partenaire commercial de la Chine en Europe.
Extracted Entity	Europe

5.2.4   Country

em/e/Country	
Definition	A country.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The entity name, based on a mention in the text and mapped to a value from an internal reference list.
Related Tag	er/Geo/Country (Disambiguation Tag)

Example1

Input Text	La Russie a hier confirmé à envoyer des renforts aux secteurs occupés comprenant les munitions et le personnel.
Extracted Entity	Russie

Example2

Input Text	L'ancien Beatles, Paul McCartney, est parti en voyage aux États-Unis avec sa compagne, Nancy Shevell.
Extracted Entity	États-Unis

5.2.5   Currency

em/e/Currency	
Definition	A reference to a currency.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The entity name, based on a mention in the text and mapped to a value from an internal reference list. **permid:** The Thomson Reuters unique ID (PermID) for this metadata tag type. Can be used when building a knowledge graph.
Related Tag	er/Currency (Disambiguation Tag)

Example1

Input Text	La correspondance de 35 lettres et cartes postales (de 1939 à 1966) était évaluée avant la vente à plus de 35.000 euros.
Extracted Entity	EUR

Example2

Input Text	l'ancien président Fidel Castro a averti que le passage de Paloma pourrait ralentir les efforts de Cuba pour se remettre des ouragans Gustav et Ike, qui ont causé 9,4 milliards de dollars de dégâts (7,3 milliards d'euros) et détruit une partie des récoltes sur l'île.
Extracted Entity	USD EUR

5.2.6 EmailAddress

em/e/EmailAddress	
Definition	An email address.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	Les demandes d'autorisation doivent être adressées par courrier électronique à publications@wto.org.
Extracted Entity	publications@wto.org

5.2.7   FaxNumber

em/e/FaxNumber	
Definition	A fax number, including the prefix and extension if they are present in the input text.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	Le Bureau Langelier du SCC, au téléphone (514) 493-0995 et au télécopieur (514) 493-3306, assure la liaison.
Extracted Entity	(514) 493-3306

5.2.8   MarketIndex

em/e/MarketIndex	
Definition	A stock market index.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The entity name, based on a mention in the text and mapped to a value from an internal reference list.
	permid: The Thomson Reuters unique ID (PermID) for this metadata tag type. Can be used when building a knowledge graph.
	ric: The Market Index Reuters Instrument Code (RIC), if it is mentioned in the text.

Example

Input Text	Le Nasdaq a glissé de 46,13 points à 1505,9, tandis que le S&P 500 a reculé de 27,85 points à 848,92.
Extracted Entity	NASDAQ 100
	S&P 500

5.2.9   NaturalFeature

em/e/NaturalFeature	
Definition	A natural feature such as a river, sea, lake, or mountain, or the name of a geographical region.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example 1

Input Text	Sur les pentes du mont Solisko, plusieurs pistes offrent d'excellentes conditions de ski alpin pour tous niveaux.
Extracted Entity	mont Solisko

Example 2

Input Text	Les pêcheurs avaient l'habitude de migrer le long des berges et vers les lacs Habaniya et Tharthar, qui se trouvent en zone sunnite.
Extracted Entity	lac Habaniya lac Tharthar

5.2.10   Organization

em/e/Organization	
Definition	The name of an organization (governmental, military, or other organization).
	Note: A business organization that generates an em/e/Company tag (that is extracted as a company) will not generate an em/e/Organization tag.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The organization name.
	nationality: The nationality of the extracted Person, Company, or Organization, if it is mentioned in the text.
	organizationtype: An organization classification. Possible values: sports, governmental military, governmental civilian, political party, N/A.
	permid: The Thomson Reuters unique ID (PermID) for this metadata tag type. Can be used when building a knowledge graph.
Related Tag	er/Organization (Disambiguation Tag)

Example

Input Text	M. Obasanjo doit s'entretenir samedi matin à Kinshasa avec les ambassadeurs des pays membres du Conseil de sécurité de l'ONU en poste en RDC.
Extracted Entity	Conseil de sécurité de l'ONU

5.2.11 Person

em/e/Person	
Definition	The name of a person. If the person is mentioned more than once in the document, the most complete mention is extracted.
Attributes	**commonname**: The commonly used person name.
	confidencelevel: A confidence score on a scale of 0 to 1. The value represents the probability that the extracted entity is indeed of the assigned type.
	For example:
	• The confidence level score for an em/e/person tag indicates the probability that the extracted person is indeed a person.
	• The confidence level score for an em/e/company tag indicates the probability that the extracted company is indeed a company.
	• The confidence level score for an em/e/pharmaceuticalDrug tag indicates the probability that the extracted pharmaceutical drug is indeed a pharmaceutical drug.
	The higher the value, the higher the probability.
	The consuming application can use this score to achieve higher accuracy results by ignoring instances with confidence scores below a specified level. Note that boosting Precision in this manner is at the expense of Recall.
	Every em/e tag that displays this attribute also has a corresponding Confidence tag.
	firstname: Most complete mention of the first name.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	lastname: Most complete mention of the last name.
	middlename: Most complete mention of the middle name.
	name: The name of the person as it appears in the text. If the person is mentioned more than once in the text, then the most complete mention is used.
	nationality: The nationality of the extracted Person, Company, or Organization, if it is mentioned in the text.
	permid: The Thomson Reuters unique ID (PermID) for this metadata tag type. Can be used when building a knowledge graph.
	persontype: A person classification. Possible values: sports, entertainment, political, economic, military, NA.
Related Tags	Confidence, RelevanceInfo, er/Person (Disambiguation Tag)

Example 1

Input Text	De son côté, le Premier secrétaire sortant François Hollande s'efforçait de dédramatiser.
Extracted Entity	François Hollande

Example 2

Input Text	Sur le terrain de West Brom, Jose Bosingwa a ouvert le score avant qu'Anelka ne trouve le chemin des filets deux fois en l'espace de 13 minutes.
Extracted Entity	Jose Bosingwa

5.2.12   PhoneNumber

em/e/PhoneNumber	
Definition	A phone number, complete with the prefix and extension if they are present in the input text.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	Les billets pour ce spectacle seront disponibles au coût de 45,50$ à 59,50$ au Grand Théâtre, par téléphone au (418) 643-8131 ou au 1 877 643-8131 ainsi que par l'entremise du réseau Billetech.
Extracted Entity	(418) 643-8131 1 877 643-8131

5.2.13   ProvinceOrState

em/e/ProvinceOrState	
Definition	A province, state, county, or other jurisdiction that is part of a country.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
Related Tag	er/Geo/ProvinceOrState (Disambiguation Tag)

Example

Input Text	Accompagné du général Jacques Grandchamp, commandant de la région de gendarmerie de Rhône-Alpes, M. Gaudin a souligné que, jusqu'à présent, aucun élément ne tendait vers autre chose qu'une disparition "accidentelle".
Extracted Entity	Rhône-Alpes

5.2.14   Region

em/e/Region	
Definition	A non-politically defined and not natural geographical region of the world (excluding continents). For example, the Middle East.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
Related Tag	er/Region (disambiguation tag)
Related Tag	er/Region

Example

Input Text	DJAVA, Géorgie - La Géorgie a lancé vendredi une offensive militaire de grande envergure pour reprendre le contrôle de l'Ossétie du Sud, dans le nord du pay.
Extracted Entity	Ossétie du Sud

5.2.15　URL

em/e/URL	
Definition	A URL or FTP address.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.

Example

Input Text	Les billets seront mis en vente samedi le 15 novembre à la billetterie de la Place des Arts, par Internet à www.pda.qc.ca.
Extracted Entity	www.pda.qc.ca

5.3   RelevanceInfo Tag

For a conceptual explanation of this tag, see Relevance Tags.

RelevanceInfo	
Definition	Indicates how centric the associated entity is to the document.
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **subject**: A hash tag generated by Open Calais. This ID points to the associated extracted entity. **Note:** In the JSON output format, because related tags are nested, the docid and subject attributes are not displayed in the RelevanceInfo tag. **relevance**: The relevance score. The higher the score, the greater the relevance of the entity to the document. Currently, the following values are supported: • 1.0 – This value is reserved for the company identified as the reporting company in a document with a predefined format such as an SEC report. • 0.8 – Entities defined as having high relevance will receive this score. • 0.5 – Entities defined as having medium relevance will receive this score. • 0.2 – Entities defined as having low relevance will receive this relevance score. • 0.0 – This value is reserved for entities identified as irrelevant to the story. For example, mentions of companies as rating agencies, reporting agencies, stock exchanges, and social applications receive a relevance score of 0. **relevancecont:** A more granular relevance score. This attribute is relevant only to Company entities. This attribute is not relevant to JSON output. • A score that is 0.8 and above indicates a high relevance. • A score of 0 indicates zero relevance. For both **relevance** and **relevancecont** scores, the three following buckets are the most indicative ones: • High relevance: score is equal to or greater than 0.8 • Zero relevance: score equals 0 • Everything else: score is greater than zero and less than 0.8 **Note:** For the highest accuracy, we recommend ranking companies based on the **relevance** score and *not* on the **relevancecont** score

RelevanceInfo

Example

Open Calais Output	*RDF...*

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
        bf3520cddfa6/Relevance/47">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/RelevanceInfo"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
        bf3520cddfa6"/>
        <c:subject rdf:resource="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-
        1ea3200d37e4"/>
        <c:relevance>0.8</c:relevance>
        <c:relevancecont>0.72</c:relevance>
    </rdf:Description>
```

N3...

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Relevance/47>
        a                       <http://s.opencalais.com/1/type/sys/RelevanceInfo> ;
        c:docId                 <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-
        816cb722a630> ;
        c:relevance     "0.8" ;
        c:relevancecont "0.72" ;
        c:subject               <http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-
        1ea3200d37e4> .
```

JSON...

```
        "relevance": 0.8
    },
```

5.4   Confidence Tag

Currently, Open Calais generates Confidence tags for the following entity and relation types: Company, Person, Pharmaceutical Drug, Bankruptcy, Deal, IPO.

The consuming application can achieve higher accuracy results by ignoring instances with confidence scores below a specified level. Note, however, that raising the specified level boosts Precision at the expense of Recall.

For a conceptual explanation of this tag, see Confidence Tags.

Confidence	
Definition	Indicates the probability that the associated e.g. extracted person or company is indeed a person or company.
Attributes	**aggregate**: The confidence score. **dblookup**: A value used internally by Open Calais to calculate the confidence score. **docid**: The unique ID of the containing document, generated by Open Calais, and formatted as a URI. **resolution**: A value used internally by Open Calais to calculate the confidence score. **statisticalfeature**: A value used internally by Open Calais to calculate the confidence score. **subject**: A hash tag generated by Open Calais. This ID points to the associated extracted entity or relation. **Note:** In the JSON output format, because related tags are nested, the docid and subject attributes are not displayed in the Confidence tag.

Confidence

Example

Open Calais Output	
	RDF...

	<pre><rdf:Description rdf:about="http://d.opencalais.com/conf/comphash-1/705cd5cf-93e1-323c-8d4e-
	        1ea3200d37e4">
	        <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/Confidence"/>
	        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
	        bf3520cddfa6"/>
	        <!--Apple-->
	        <c:subject rdf:resource="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-
	        1ea3200d37e4"/>
	        <c:statisticalfeature>0.997</c:statisticalfeature>
	        <c:dblookup>0.0</c:dblookup>
	        <c:resolution>0.9928677</c:resolution>
	        <c:aggregate>0.996</c:aggregate>
	    </rdf:Description></pre>

	N3...

	<pre><http://d.opencalais.com/conf/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4>
	        a <http://s.opencalais.com/1/type/tag/Confidence> ;
	        c:aggregate "0.996" ;
	        c:dblookup "0.0" ;
	        c:docId <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-
	        816cb722a630> ;
	        c:resolution "0.9928677" ;
	        c:statisticalfeature "0.997" ;
	        c:subject <http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-
	        1ea3200d37e4> .</pre>

	JSON...

	<pre> "confidence": {
	            "statisticalfeature": "0.997",
	            "dblookup": "0.0",
	            "resolution": "0.9928677",
	            "aggregate": "0.996"
	        },</pre>

5.5   Disambiguation Tags

Currently, Calais supports Disambiguation for the following entity and relation types: City, Company, Continent, Country, Currency, Organization, Person, ProvinceOrState, Region, TopmostPublicParentCompany.

Click a link for a description of the tag and its attributes.

For a conceptual explanation of this tag, see Disambiguation Tags.

Note: It is important to note that most attributes are optional; a tag can be extracted with some but not all of its attributes.

5.5.1   City

er/Geo/City	
Definition	This tag is generated when an extracted entity of the type City is successfully mapped to a city in the Thomson Reuters Classification Schema (TRCS).
Attributes	**containedbycountry**: The name of the country in which the geographical entity is located.
	containedbystate: The name of the state in which the geographical entity is located.
	docid: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	latitude: The latitude of the geographical entity, as defined in Freebase.
	longitude: The longitude of the geographical entity, as defined in Freebase.
	name: The official name as it appears in the Thomson Reuters dataset.
	permid: The Thomson Reuters unique ID (PermID) for this geographical entity.
	rcscode: The code of this geographical entity, as defined in the Thomson Reuters Coding Schema (TRCS).
	shortname: A name commonly used to refer to the geographical entity, as defined in Freebase.
	subject: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/City

5.5.2   Company

er/Company	
Definition	This tag is generated when an extracted entity of the type Company is mapped to a company in the Thomson Reuters dataset.
	Note that the *openpermid* attribute gives you direct access to high quality, curated Thomson Reuters company data. (In JSON output, the attribute name is *id*.)
Attributes	**commonname**: The commonly used name, as it appears in the Thomson Reuters dataset.
	docid: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	id: See the *openpermid* attribute description, below.
	ispublic: Indicates whether the company is public (true) or private (false).
	legacyid: The link to the company as it would have been displayed in the legacy Open Calais er/company tag.
	This attribute is relevant to users who upgraded from the legacy Open Calais to the new Open Calais.
	The purpose of this attribute is to enable you to map the old links, which are no longer supported, to the corresponding new links.
	This attribute is not relevant to the JSON output format.
	An example of an old company link in the legacy Open Calais er/company tag (RDF output):
	``` <rdf:Description rdf:about="http://d.opencalais.com/er/company/ralg-tr1r/39d9a86f-6241-3688-b05b-4595eeaba2d1"> ```
	The company link and the legacyid attribute as they appear in the new Open Calais er/company tag (RDF output):
	``` <rdf:Description rdf:about="http://d.opencalais.com/er/company/ralg-oa/5000608903"> <c:legacyid rdf:resource="http://d.opencalais.com/er/company/ralg-tr1r/39d9a86f-6241-3688-b05b-4595eeaba2d1"/> ```
	The company link and the legacyid attribute as they appear in the new Open Calais er/company tag (N3 output):
	``` <http://d.opencalais.com/er/company/ralg-oa/5000608903> c:legacyid < http://d.opencalais.com/er/company/ralg-tr1r/39d9a86f-6241-3688-b05b-4595eeaba2d1> ; ```
	name: The official name as it appears in the Thomson Reuters dataset.
	openpermid: A direct link to the relevant company page on the Open PermID website (https://permid.org). This link gives you direct access to high quality, curated Thomson Reuters company data. (In the JSON output, the attribute name is **id**.)
	permid: The Thomson Reuters unique ID (PermID) for this company.
	primaryric: The company's primary Reuters Instrument code (RIC) as it appears in the Thomson Reuters dataset.
	score: A score representing the probability that the extracted company was accurately mapped to the corresponding company in the Thomson Reuters dataset.
	Scores range from 0 - 1; the higher the score, the higher the probability. The consuming application can use this score to achieve higher accuracy results.
	subject: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
	ticker: The company stock symbol.
Related Tags	em/e/Company er/TopmostPublicParentCompany

5.5.3   Continent

er/Continent	
Definition	This tag is generated when an extracted entity of the type Continent is successfully mapped to a continent in the Thomson Reuters Classification Schema (TRCS).
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **name**: The official name as it appears in the Thomson Reuters dataset. **permid**: The Thomson Reuters unique ID (PermID) for this geographical entity. **rcscode**: The code of this geographical entity, as defined in the Thomson Reuters Coding Schema (TRCS). **subject**: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Continent

5.5.4 Country

er/Geo/Country	
Definition	This tag is generated when an extracted entity of the type Country is successfully mapped to a country in the Thomson Reuters Classification Schema (TRCS).
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **latitude**: The latitude of the geographical entity, as defined in Freebase. **longitude**: The longitude of the geographical entity, as defined in Freebase. **name**: The official name as it appears in the Thomson Reuters dataset. **permid**: The Thomson Reuters unique ID (PermID) for this geographical entity. **rcscode**: The code of this geographical entity, as defined in the Thomson Reuters Coding Schema (TRCS). **shortname**: A name commonly used to refer to the geographical entity, as defined in Freebase. **subject**: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Country

5.5.5　Currency

er/Currency	
Definition	This tag is generated when an extracted entity of the type Currency is successfully mapped to a currency in the Thomson Reuters Classification Schema (TRCS).
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **name**: The official name as it appears in the Thomson Reuters dataset. **permid**: The Thomson Reuters unique ID (PermID) for this currency. **rcscode**: The currency code, as defined in the Thomson Reuters Coding Schema (TRCS). **subject**: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Currency

5.5.6　　Organization

er/Organization	
Definition	This tag is generated when an extracted entity of the type Organization is successfully mapped to an organization in the Thomson Reuters Organization Authority dataset.
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **name**: The official name as it appears in the Thomson Reuters dataset. **permid**: The Thomson Reuters unique ID (PermID) for this organization. **rcscode**: The organization code, as defined in the Thomson Reuters Coding Schema (TRCS). **subject**: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Organization

5.5.7 Person

er/Person	
Definition	This tag is generated when an extracted entity of the type Person is successfully mapped to a person in the Thomson Reuters People Authority dataset. Currently this tag is only generated for executives.
Attributes	**commonname**: The commonly used name, as it appears in the Thomson Reuters dataset.
	docid: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	name: The official name as it appears in the Thomson Reuters dataset.
	officerid: A unique ID (as defined in the Thomson Reuters Officers and Directors dataset) that specifies a position held (currently or in the past) by the person, in a particular company. The er/person tag may include multiple officerid attributes.
	openpermid: A direct link to the relevant person page on the Open PermID website (https://permid.org). This link gives you direct access to high quality, curated Thomson Reuters people data. (In the JSON output, the attribute name is **id**.)
	paid: The Thomson Reuters unique ID (from the People Authority) for this person.
	personid: The Person ID as defined in the Thomson Reuters Officers and Directors dataset.
	score: A score representing the probability that the extracted person was accurately mapped to the corresponding person in the Thomson Reuters dataset.
	Scores range from 0 - 1; the higher the score, the higher the probability. The consuming application can use this score to achieve higher accuracy results.
	subject: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Person

5.5.8 ProvinceOrState

er/Geo/ProvinceOrState	
Definition	This tag is generated when an extracted entity of the type ProvinceOrState is successfully mapped to a province or state in the Thomson Reuters Classification Schema (TRCS).
Attributes	**containedbycountry**: The name of the country in which the geographical entity is located.
	containedbystate: The name of the state in which the geographical entity is located.
	docid: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	latitude: The latitude of the geographical entity, as defined in Freebase.
	longitude: The longitude of the geographical entity, as defined in Freebase.
	name: The official name as it appears in the Thomson Reuters dataset.
	permid: The Thomson Reuters unique ID (PermID) for this geographical entity.
	rcscode: The code of this geographical entity, as defined in the Thomson Reuters Coding Schema (TRCS).
	shortname: A name commonly used to refer to the geographical entity, as defined in Freebase.
	subject: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/ProvinceOrState

5.5.9   Region

er/Region	
Definition	This tag is generated when an extracted entity of the type Region is successfully mapped to a region in the Thomson Reuters Classification Schema (TRCS).
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **name**: The official name as it appears in the Thomson Reuters dataset. **permid**: The Thomson Reuters unique ID (PermID) for this region. **rcscode**: The code of this region, as defined in the Thomson Reuters Coding Schema (TRCS). **subject**: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Region

5.5.10   TopmostPublicParentCompany

er/TopmostPublicParentCompany	
Definition	This tag is generated when an extracted entity of the type Company is mapped to a company in the Thomson Reuters dataset that also has a publicly traded parent company.
	Note that the *openpermid* attribute gives you direct access to high quality, curated Thomson Reuters company data. (In JSON output, the attribute name is *id*.)
	This metadata type is available upon subscription to Thomson Reuters Intelligent Tagging (TRIT). For more information please contact us at <u>questions@opencalais.com</u>.
Attributes	**commonname**: The commonly used name, as it appears in the Thomson Reuters dataset.
	docid: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	id: See the openpermid attribute description, below.
	ispublic: Indicates whether the company is public (true) or private (false).
	legacyid: The link to the company as it would have been displayed in the legacy Open Calais output.
	This attribute is relevant to users who upgraded from the legacy Open Calais to the new Open Calais.
	The purpose of this attribute is to enable you to map the old links, which are no longer supported, to the corresponding new links.
	This attribute is not relevant to the JSON output format.
	name: The official name as it appears in the Thomson Reuters dataset.
	openpermid: A direct link to the relevant company page on the Open PermID website (https://permid.org). This link gives you direct access to high quality, curated Thomson Reuters company data. (In the JSON output, the attribute name is **id**.).)
	permid: The Thomson Reuters unique ID (PermID) for this company.
	primaryric: The company's primary Reuters Instrument code (RIC) as it appears in the Thomson Reuters dataset.
	subject: The hash tag generated by Open Calais and assigned to the extracted child company. This ID points to the relevant em/e/company and related tags within the document. If the parent company has multiple extracted child companies within the document, additional instances of the subject attribute appear in the tag. Note that this entity is not relevant to the JSON output format.
	ticker: The company stock symbol.
Related Tag	em/e/Company

5.6   Industry Tag

This functionality is available upon subscription to Thomson Reuters Intelligent Tagging (TRIT). For more information please contact us at questions@opencalais.com.

For a conceptual explanation, see Industry Tags.

Industry	
Definition	Identifies the industries related to the companies mentioned in the document.
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The name of the industry. **permid**: The Thomson Reuters unique ID (PermID) for this Industry. **rcscode**: The industry code, as defined in the Thomson Reuters Coding Schema (TRCS). **relevance**: Indicates how relevant the industry is to the story. Values range from 0 to 1. The higher the score, the higher the relevance. **trbccode**: The industry code, as defined in the Thomson Reuters Business Classification (TRBC) taxonomy.

Industry

Example

Open Calais
Output
	RDF...
	
	``
	<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
	bf3520cddfa6/Industry/5">
	    <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/Industry"/>
	    <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
	bf3520cddfa6"/>
	    <c:forenduserdisplay>false</c:forenduserdisplay>
	    <c:name>Automobiles & Multi Utility Vehicles</c:name>
	    <c:rcscode>B:1294</c:rcscode>
	    <c:permid>4294951707</c:permid>
	    <c:relevance>0.500</c:relevance>
	</rdf:Description>
	``
	
	N3...
	
	``
	<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Industry/5>
	    a <http://s.opencalais.com/1/type/tag/Industry> ;
	    c:docId <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-
	816cb722a630> ;
	    c:forenduserdisplay "false" ;
	    c:name "Automobiles & Multi Utility Vehicles" ;
	    c:permid "4294951707" ;
	    c:rcscode "B:1294" ;
	    c:relevance "0.500" .
	``
	
	JSON...
	
	``
	"http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Industry/5": {
	    "_typeGroup": "industry",
	    "forenduserdisplay": "false",
	    "name": "Automobiles & Multi Utility Vehicles",
	    "rcscode": "B:1294",
	    "permid": "4294951707",
	    "relevance": 0.5
	},
	``

Chapter 6   Spanish Semantic Metadata Tags

Open Calais is able to extract a rich set of metadata tags from Spanish input text:

- InstanceInfo Tags

- The following Entity Markup (em/e) tags: City, Company, Continent, Country, Currency, EmailAddress, FaxNumber, MarketIndex, NaturalFeature, Organization, Person, PhoneNumber, ProvinceOrState, Region, URL.

- RelevanceInfo Tags

- Confidence Tags

- The following Disambiguation (er) tags: City, Company, Continent, Country, Currency, Organization, Person, ProvinceOrState, Region, TopmostPublicParentCompany.

- Industry Tags

(Currently, Social Tags and Topic Tags are not extracted from Spanish input text.)

For a conceptual overview of the Open Calais Semantic Metadata Tags, see How Does Open Calais Work?

6.1   InstanceInfo Tag

For a conceptual explanation of this tag, see Instance Tags.

InstanceInfo	
Definition	Describes a mention of an Open Calais type found in the text.
Attributes	**detection**: The text string in which the mention was identified.
	docid: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	exact: The mention.
	length: The length (in characters) of the mention.
	offset: Offset of the mention (in characters) from the beginning of the document.
	prefix: The portion of the text string that precedes the mention.
	subject: A hash tag generated by Open Calais. This ID points to the associated extracted entity or relation.
	suffix: The portion of the text string that follows the mention.
	The tag also includes a comment that indicates the type of entity or relation extracted from this instance, and its main attribute values.
	Note: In the JSON output format, because related tags are nested, the docid and subject attributes are not displayed in the Instance tag.

Example:

Input Text	Las cenizas de los Roddenberry serán lanzadas al espacio dentro de un año y medio, según sus deseos, dijo el lunes la compañía de vuelos espaciales Celestis Inc. Majel Barrett Roddenberry.
| Instance Tag | ```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/081a832b-3ebf-313a-972b-
      6e2c9e041ca9/Instance/1">
   <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/InstanceInfo"/>
   <c:docId rdf:resource="http://d.opencalais.com/dochash-1/081a832b-3ebf-313a-972b-
      6e2c9e041ca9"/>
   <c:subject rdf:resource="http://d.opencalais.com/comphash-1/befa8b2c-d819-3048-aea3-
      0b57798ce146"/>
   <!--Company: Celestis Inc.; -->
   <c:detection>[dijo el lunes la compañía de vuelos espaciales ]Celestis Inc.[ Majel Barrett
      Roddenberry&lt;/Body&gt;]</c:detection>
      <c:prefix>dijo el lunes la compañía de vuelos espaciales </c:prefix>
      <c:exact>Celestis Inc.</c:exact>
      <c:suffix> Majel Barrett Roddenberry&lt;/Body&gt;</c:suffix>
      <c:offset>221</c:offset>
</rdf:Description>
``` |

6.2 Entity Markup Tags

Click a link for a description of the entity type and its attributes.

City, Company, Continent, Country, Currency, EmailAddress, FaxNumber, MarketIndex, NaturalFeature, Organization, Person, PhoneNumber, ProvinceOrState, Region, URL

Note: It is important to note that most attributes are optional; a tag can be extracted with some but not all of its attributes.

For a conceptual explanation, see Entity Markup Tags.

6.2.1 City

em/e/City	
Definition	The name of a city or other municipality. Districts or neighborhoods within a city (e.g. Soho, London) are not extracted.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
Related Tag	er/Geo/City (Disambiguation Tag)

Example1

Input Text	La alta costura de París piensa en grande.
Extracted Entity	París

Example2

Input Text	los ataques del 11 de septiembre en Nueva York.
Extracted Entity	Nueva York

6.2.2 Company

em/e/Company	
Definition	A full or partial company name. The term "Company" refers to any business organization, including newspapers, media companies, law firms, etc.
Attributes	**confidencelevel**: A confidence score on a scale of 0 to 1. The value represents the probability that the extracted entity is indeed of the assigned type.
	For example:
	• The confidence level score for an em/e/person tag indicates the probability that the extracted person is indeed a person.
	• The confidence level score for an em/e/company tag indicates the probability that the extracted company is indeed a company.
	• The confidence level score for an em/e/pharmaceuticalDrug tag indicates the probability that the extracted pharmaceutical drug is indeed a pharmaceutical drug.
	The higher the value, the higher the probability.
	The consuming application can use this score to achieve higher accuracy results by ignoring instances with confidence scores below a specified level. Note that boosting Precision in this manner is at the expense of Recall.
	Every em/e tag that displays this attribute also has a corresponding Confidence tag.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	The related confidence score determines if the forenduserdisplay value is true or false.
	inlineric: The company Reuters Instrument Code (RIC), if it is mentioned in the text.
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
	nationality: The nationality of the extracted Person, Company, or Organization, if it is mentioned in the text.
Related Tags	Confidence, RelevanceInfo, er/Company (Disambiguation Tag)

Example1

Input Text	Las cenizas de los Roddenberry serán lanzadas al espacio dentro de un año y medio, según sus deseos, dijo el lunes la compañía de vuelos espaciales Celestis Inc. Majel Barrett Roddenberry.
Extracted Entity	Celestis Inc.

Example2

Input Text	Iberdrola subió un 1,5 por ciento a 6,10 euros.
Extracted Entity	Iberdrola

6.2.3 Continent

em/e/Continent	
Definition	A continent.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The entity name, based on a mention in the text and mapped to a value from an internal reference list.
Related Tag	er/Continent (Disambiguation Tag)

Example 1

Input Text	La Antártida se está calentando, no enfriando, según un estudio
Extracted Entity	Antártida

Example 2

Input Text	Jackson ha pospuesto seis apariciones en Norteamérica.
Extracted Entity	Norteamérica

6.2.4   Country

em/e/Country	
Definition	A country.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The entity name, based on a mention in the text and mapped to a value from an internal reference list.
Related Tag	er/Geo/Country (Disambiguation Tag)

Example

Input Text	La policía de Bahamas investiga la acusación de extorsión a Travolta.
Extracted Entity	Bahamas

6.2.5　Currency

em/e/Currency	
Definition	A reference to a currency.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The entity name, based on a mention in the text and mapped to a value from an internal reference list.
	permid: The Thomson Reuters unique ID (PermID) for this metadata tag type. Can be used when building a knowledge graph.
Related Tag	er/Currency (Disambiguation Tag)

Example1

Input Text	Para poner las cosas en perspectiva, al comprar en grandes cantidades, los fabricantes de computadoras normalmente pagan $50-$60 por cada copia de Windows Vistas, mientras que la versión Starter está alrededor de $30.
Extracted Entity	USD

6.2.6   EmailAddress

em/e/EmailAddress	
Definition	An email address.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
Example	
Input Text	La Escuela Comunitaria de Austin P.O. Box 6176 Austin, TX 78762-6176 oficina: 512-554-8930 fax: 866-868-9973 info@austincommunityschool.org
Extracted Entity	info@austincommunityschool.org

6.2.7   FaxNumber

em/e/FaxNumber	
Definition	A fax number, including the prefix and extension if they are present in the input text.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
Example	
Input Text	Tamarindo - Ciudad
	100 Metros Norte y 25 Metros Oeste de Nunciatura
	Tel. +(506) 2653-0012 Fax: +(506) 2653-0012
Extracted Entity	+(506) 2653-0012

6.2.8   MarketIndex

em/e/MarketIndex	
Definition	A stock market index.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The entity name, based on a mention in the text and mapped to a value from an internal reference list. **permid:** The Thomson Reuters unique ID (PermID) for this metadata tag type. Can be used when building a knowledge graph. **ric**: The Market Index Reuters Instrument Code (RIC), if it is mentioned in the text.
Example	
Input Text	A las 0844 GMT el índice paneuropeo FTSEurofirst 300 subîa un 0,3 por ciento.
Extracted Entity	FTSEurofirst 300

6.2.9   NaturalFeature

em/e/NaturalFeature	
Definition	A natural feature such as a river, sea, lake, or mountain, or the name of a geographical region.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
Example 1	
Input Text	Por primera vez en España, organizamos una expedición al corazón del Himalaya hindú, al macizo de Garhwal.
Extracted Entity	Himilaya Garhwal

Example 2

Input Text	Caminar sobre el milenario hielo azul, navegar por el lago Argentino para ver de cerca otros gigantescos ríos de hielo.
Extracted Entity	lago Argentino

6.2.10   Organization

em/e/Organization	
Definition	The name of an organization (governmental, military, or other organization).
	Note: A business organization that generates an em/e/Company tag (that is extracted as a company) will not generate an em/e/Organization tag.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name:** The organization name. **nationality**: The nationality of the extracted Person, Company, or Organization, if it is mentioned in the text. **organizationtype**: An organization classification. Possible values: sports, governmental military, governmental civilian, political party, N/A. **permid:** The Thomson Reuters unique ID (PermID) for this metadata tag type. Can be used when building a knowledge graph.
Related Tag	er/Organization (Disambiguation Tag)

Example 1

Input Text	Vegara dijo que la tasa de paro no llegaría al 18,7 por ciento pronosticado por la Comisión Europea en 2010
Extracted Entity	la Comisión Europea

Example 2

Input Text	cada vez que la Casa Blanca pasa de un partido a otro.
Extracted Entity	Casa Blanca

6.2.11   Person

em/e/Person	
Definition	The name of a person. If the person is mentioned more than once in the document, the most complete mention is extracted.
Attributes	**commonname**: The commonly used person name. **confidencelevel**: A confidence score on a scale of 0 to 1. The value represents the probability that the extracted entity is indeed of the assigned type. For example: • The confidence level score for an em/e/person tag indicates the probability that the extracted person is indeed a person. • The confidence level score for an em/e/company tag indicates the probability that the extracted company is indeed a company. • The confidence level score for an em/e/pharmaceuticalDrug tag indicates the probability that the extracted pharmaceutical drug is indeed a pharmaceutical drug. The higher the value, the higher the probability. The consuming application can use this score to achieve higher accuracy results by ignoring instances with confidence scores below a specified level. Note that boosting Precision in this manner is at the expense of Recall. Every em/e tag that displays this attribute also has a corresponding Confidence tag. **firstname**: Most complete mention of the first name. **forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **lastname**: Most complete mention of the last name. **middlename**: Most complete mention of the middle name. **name**: The name of the person as it appears in the text. If the person is mentioned more than once in the text, then the most complete mention is used. **nationality**: The nationality of the extracted Person, Company, or Organization, if it is mentioned in the text. **permid:** The Thomson Reuters unique ID (PermID) for this metadata tag type. Can be used when building a knowledge graph. **persontype**: A person classification. Possible values: sports, entertainment, political, economic, military, NA.
Related Tags	Confidence, RelevanceInfo, er/Person (Disambiguation Tag)

Example 1

Input Text	Las autoridades de Bahamas están investigando un supuesto intento de extorsión contra el actor John Travolta.
Extracted Entity	John Travolta

Example 2

Input Text	El comisario de policía Reginald Ferguson dijo que sus agentes estaban investigando una queja de extorsión.
Extracted Entity	Reginald Ferguson

6.2.12   PhoneNumber

em/e/PhoneNumber	
Definition	A phone number, complete with the prefix and extension if they are present in the input text.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
Example	
Input Text	La Escuela Comunitaria de Austin
	P.O. Box 6176
	Austin, TX 78762-6176
	oficina: 512-554-8930 fax: 866-868-9973
	info@austincommunityschool.org
	http://www.campo21.com/
Extracted Entity	512-554-8930

6.2.13   ProvinceOrState

em/e/ProvinceOrState	
Definition	A province, state, county, or other jurisdiction that is part of a country.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
Related Tag	er/Geo/ProvinceOrState (Disambiguation Tag)

Example 1

Input Text	Interior confirmó la detención de otras tres personas en Vizcaya, dos en Guipúzcoa y una en Álava.
Extracted Entity	Vizcaya Guipúzcoa Álava

Example 2

Input Text	por la noche en la provincia de La Coruña.
Extracted Entity	La Coruña

6.2.14   Region

em/e/Region	
Definition	A non-politically defined and not natural geographical region of the world (excluding continents). For example, the Middle East.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
Related Tag	er/Region (Disambiguation Tag)

Example 1

Input Text	Dos personas han muerto en España debido al temporal de viento que afecta al norte de la Península Ibérica.
Extracted Entity	Península Ibérica

Example 2

Input Text	en el suroeste de Francia se registraron fuertes vientos.
Extracted Entity	el suroeste de Francia

6.2.15　URL

em/e/URL	
Definition	A URL or FTP address.
Attributes	**forenduserdisplay**: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). **name**: The text string from the input document that represents the entity. When there are multiple mentions that refer to the same entity, the most complete mention is used.
Example	
Input Text	Visite el nuevo portal http://www.carilo.com.ar toda la información de esta ciudad, servicios, alojamientos, inmobiliarias, etc.
Extracted Entity	http://www.carilo.com.ar

6.3 RelevanceInfo Tag

For a conceptual explanation of this tag, see Relevance Tags.

RelevanceInfo	
Definition	Indicates how centric the associated entity is to the document.
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **subject**: A hash tag generated by Open Calais. This ID points to the associated extracted entity. --- **Note:** In the JSON output format, because related tags are nested, the docid and subject attributes are not displayed in the RelevanceInfo tag. --- **relevance**: The relevance score. The higher the score, the greater the relevance of the entity to the document. Currently, the following values are supported: • 1.0 – This value is reserved for the company identified as the reporting company in a document with a predefined format such as an SEC report. • 0.8 – Entities defined as having high relevance will receive this score. • 0.5 – Entities defined as having medium relevance will receive this score. • 0.2 – Entities defined as having low relevance will receive this relevance score. • 0.0 – This value is reserved for entities identified as irrelevant to the story. For example, mentions of companies as rating agencies, reporting agencies, stock exchanges, and social applications receive a relevance score of 0. **relevancecont:** A more granular relevance score. This attribute is relevant only to Company entities. This attribute is not relevant to JSON output. • A score that is 0.8 and above indicates a high relevance. • A score of 0 indicates zero relevance. For both **relevance** and **relevancecont** scores, the three following buckets are the most indicative ones: • High relevance: score is equal to or greater than 0.8 • Zero relevance: score equals 0 • Everything else: score is greater than zero and less than 0.8 --- **Note:** For the highest accuracy, we recommend ranking companies based on the **relevance** score and *not* on the **relevancecont** score ---

RelevanceInfo

Example

Open Calais Output	*RDF...*
	```<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6/Relevance/47">```
	``` <rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/RelevanceInfo"/>```
	``` <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-bf3520cddfa6"/>```
	``` <c:subject rdf:resource="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4"/>```
	``` <c:relevance>0.8</c:relevance>```
	``` <c:relevancecont>0.72</c:relevance>```
	```</rdf:Description>```
	N3...
	```<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Relevance/47>```
	``` a <http://s.opencalais.com/1/type/sys/RelevanceInfo> ;```
	``` c:docId <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630> ;```
	``` c:relevance "0.8" ;```
	``` c:relevancecont "0.72" ;```
	``` c:subject <http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4> .```
	JSON...
	``` "relevance": 0.8```
	``` },```

6.4　Confidence Tag

Currently, Open Calais generates Confidence tags for the following entity and relation types: Company, Person, Pharmaceutical Drug, Bankruptcy, Deal, IPO.

The consuming application can achieve higher accuracy results by ignoring instances with confidence scores below a specified level. Note, however, that raising the specified level boosts Precision at the expense of Recall.

For a conceptual explanation of this tag, see Confidence Tags.

Confidence	
Definition	Indicates the probability that the associated e.g. extracted person or company is indeed a person or company.
Attributes	**aggregate**: The confidence score.
	dblookup: A value used internally by Open Calais to calculate the confidence score.
	docid: The unique ID of the containing document, generated by Open Calais, and formatted as a URI.
	resolution: A value used internally by Open Calais to calculate the confidence score.
	statisticalfeature: A value used internally by Open Calais to calculate the confidence score.
	subject: A hash tag generated by Open Calais. This ID points to the associated extracted entity or relation.
	Note:　In the JSON output format, because related tags are nested, the docid and subject attributes are not displayed in the Confidence tag.

Confidence

Example

Open Calais Output	*RDF...*
| | ```
<rdf:Description rdf:about="http://d.opencalais.com/conf/comphash-1/705cd5cf-93e1-323c-8d4e-
        1ea3200d37e4">
        <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/Confidence"/>
        <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
        bf3520cddfa6"/>
        <!--Apple-->
        <c:subject rdf:resource="http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-
        1ea3200d37e4"/>
        <c:statisticalfeature>0.997</c:statisticalfeature>
        <c:dblookup>0.0</c:dblookup>
        <c:resolution>0.9928677</c:resolution>
        <c:aggregate>0.996</c:aggregate>
    </rdf:Description>
``` |
| | *N3...* |
| | ```
<http://d.opencalais.com/conf/comphash-1/705cd5cf-93e1-323c-8d4e-1ea3200d37e4>
        a                       <http://s.opencalais.com/1/type/tag/Confidence> ;
        c:aggregate             "0.996" ;
        c:dblookup              "0.0" ;
        c:docId                 <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-
816cb722a630> ;
        c:resolution            "0.9928677" ;
        c:statisticalfeature    "0.997" ;
        c:subject               <http://d.opencalais.com/comphash-1/705cd5cf-93e1-323c-8d4e-
1ea3200d37e4> .
``` |
| | *JSON...* |
| | ```
        "confidence": {
            "statisticalfeature": "0.997",
            "dblookup": "0.0",
            "resolution": "0.9928677",
            "aggregate": "0.996"
        },
``` |

6.5   Disambiguation Tags

Currently, Calais supports Disambiguation for the following entity and relation types: City, Company, Continent, Country, Currency, Organization, Person, ProvinceOrState, Region, TopmostPublicParentCompany.

Click a link for a description of the tag and its attributes.

For a conceptual explanation of this tag, see Disambiguation Tags.

Note: It is important to note that most attributes are optional; a tag can be extracted with some but not all of its attributes.

6.5.1   City

er/Geo/City	
Definition	This tag is generated when an extracted entity of the type City is successfully mapped to a city in the Thomson Reuters Classification Schema (TRCS).
Attributes	**containedbycountry**: The name of the country in which the geographical entity is located.
	containedbystate: The name of the state in which the geographical entity is located.
	docid: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	latitude: The latitude of the geographical entity, as defined in Freebase.
	longitude: The longitude of the geographical entity, as defined in Freebase.
	name: The official name as it appears in the Thomson Reuters dataset.
	permid: The Thomson Reuters unique ID (PermID) for this geographical entity.
	rcscode: The code of this geographical entity, as defined in the Thomson Reuters Coding Schema (TRCS).
	shortname: A name commonly used to refer to the geographical entity, as defined in Freebase.
	subject: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/City

6.5.2　Company

er/Company	
Definition	This tag is generated when an extracted entity of the type Company is mapped to a company in the Thomson Reuters dataset. Note that the *openpermid* attribute gives you direct access to high quality, curated Thomson Reuters company data. (In JSON output, the attribute name is *id*.)

Attributes	**commonname**: The commonly used name, as it appears in the Thomson Reuters dataset.
	
	 docid: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	
	 id: See the *openpermid* attribute description, below.
	
	 ispublic: Indicates whether the company is public (true) or private (false).
	
	 legacyid: The link to the company as it would have been displayed in the legacy Open Calais er/company tag.
	
	 This attribute is relevant to users who upgraded from the legacy Open Calais to the new Open Calais.
	
	 The purpose of this attribute is to enable you to map the old links, which are no longer supported, to the corresponding new links.
	
	 This attribute is not relevant to the JSON output format.
	
	 An example of an old company link in the legacy Open Calais er/company tag (RDF output):
	
	 ```<rdf:Description rdf:about="http://d.opencalais.com/er/company/ralg-tr1r/39d9a86f-6241-3688-b05b-4595eeaba2d1">```
	
	 The company link and the legacyid attribute as they appear in the new Open Calais er/company tag (RDF output):
	
	 ```<rdf:Description rdf:about="http://d.opencalais.com/er/company/ralg-oa/5000608903"> <c:legacyid rdf:resource="http://d.opencalais.com/er/company/ralg-tr1r/39d9a86f-6241-3688-b05b-4595eeaba2d1"/>```
	
	 The company link and the legacyid attribute as they appear in the new Open Calais er/company tag (N3 output):
	
	 ```<http://d.opencalais.com/er/company/ralg-oa/5000608903> c:legacyid < http://d.opencalais.com/er/company/ralg-tr1r/39d9a86f-6241-3688-b05b-4595eeaba2d1> ;```
	
	 name: The official name as it appears in the Thomson Reuters dataset.
	
	 openpermid: A direct link to the relevant company page on the Open PermID website (https://permid.org). This link gives you direct access to high quality, curated Thomson Reuters company data. (In the JSON output, the attribute name is **id**.)
	
	 permid: The Thomson Reuters unique ID (PermID) for this company.
	
	 primaryric: The company's primary Reuters Instrument code (RIC) as it appears in the Thomson Reuters dataset.
	
	 score: A score representing the probability that the extracted company was accurately mapped to the corresponding company in the Thomson Reuters dataset.
	
	 Scores range from 0 - 1; the higher the score, the higher the probability. The consuming application can use this score to achieve higher accuracy results.
	
	 subject: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
	
	 ticker: The company stock symbol.

Related Tags	em/e/Company er/TopmostPublicParentCompany

6.5.3   Continent

er/Continent	
Definition	This tag is generated when an extracted entity of the type Continent is successfully mapped to a continent in the Thomson Reuters Classification Schema (TRCS).
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **name**: The official name as it appears in the Thomson Reuters dataset. **permid**: The Thomson Reuters unique ID (PermID) for this geographical entity. **rcscode**: The code of this geographical entity, as defined in the Thomson Reuters Coding Schema (TRCS). **subject**: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Continent

6.5.4   Country

er/Geo/Country	
Definition	This tag is generated when an extracted entity of the type Country is successfully mapped to a country in the Thomson Reuters Classification Schema (TRCS).
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **latitude**: The latitude of the geographical entity, as defined in Freebase. **longitude**: The longitude of the geographical entity, as defined in Freebase. **name**: The official name as it appears in the Thomson Reuters dataset. **permid**: The Thomson Reuters unique ID (PermID) for this geographical entity. **rcscode**: The code of this geographical entity, as defined in the Thomson Reuters Coding Schema (TRCS). **shortname**: A name commonly used to refer to the geographical entity, as defined in Freebase. **subject**: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Country

6.5.5   Currency

er/Currency	
Definition	This tag is generated when an extracted entity of the type Currency is successfully mapped to a currency in the Thomson Reuters Classification Schema (TRCS).
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **name**: The official name as it appears in the Thomson Reuters dataset. **permid**: The Thomson Reuters unique ID (PermID) for this currency. **rcscode**: The currency code, as defined in the Thomson Reuters Coding Schema (TRCS). **subject**: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Currency

6.5.6   Organization

er/Organization	
Definition	This tag is generated when an extracted entity of the type Organization is successfully mapped to an organization in the Thomson Reuters Organization Authority dataset.
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **name**: The official name as it appears in the Thomson Reuters dataset. **permid**: The Thomson Reuters unique ID (PermID) for this organization. **rcscode**: The organization code, as defined in the Thomson Reuters Coding Schema (TRCS). **subject**: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Organization

6.5.7   Person

er/Person	
Definition	This tag is generated when an extracted entity of the type Person is successfully mapped to a person in the Thomson Reuters People Authority dataset. Currently this tag is only generated for executives.
Attributes	**commonname**: The commonly used name, as it appears in the Thomson Reuters dataset.
	docid: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	name: The official name as it appears in the Thomson Reuters dataset.
	officerid: A unique ID (as defined in the Thomson Reuters Officers and Directors dataset) that specifies a position held (currently or in the past) by the person, in a particular company. The er/person tag may include multiple officerid attributes.
	openpermid: A direct link to the relevant person page on the Open PermID website (https://permid.org). This link gives you direct access to high quality, curated Thomson Reuters people data. (In the JSON output, the attribute name is **id**.)
	paid: The Thomson Reuters unique ID (from the People Authority) for this person.
	personid: The Person ID as defined in the Thomson Reuters Officers and Directors dataset.
	score: A score representing the probability that the extracted person was accurately mapped to the corresponding person in the Thomson Reuters dataset.
	Scores range from 0 - 1; the higher the score, the higher the probability. The consuming application can use this score to achieve higher accuracy results.
	subject: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Person

6.5.8   ProvinceOrState

er/Geo/ProvinceOrState	
Definition	This tag is generated when an extracted entity of the type ProvinceOrState is successfully mapped to a province or state in the Thomson Reuters Classification Schema (TRCS).
Attributes	**containedbycountry**: The name of the country in which the geographical entity is located.
	containedbystate: The name of the state in which the geographical entity is located.
	docid: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	latitude: The latitude of the geographical entity, as defined in Freebase.
	longitude: The longitude of the geographical entity, as defined in Freebase.
	name: The official name as it appears in the Thomson Reuters dataset.
	permid: The Thomson Reuters unique ID (PermID) for this geographical entity.
	rcscode: The code of this geographical entity, as defined in the Thomson Reuters Coding Schema (TRCS).
	shortname: A name commonly used to refer to the geographical entity, as defined in Freebase.
	subject: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/ProvinceOrState

6.5.9   Region

er/Region	
Definition	This tag is generated when an extracted entity of the type Region is successfully mapped to a region in the Thomson Reuters Classification Schema (TRCS).
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **name**: The official name as it appears in the Thomson Reuters dataset. **permid**: The Thomson Reuters unique ID (PermID) for this region. **rcscode**: The code of this region, as defined in the Thomson Reuters Coding Schema (TRCS). **subject**: The hash tag generated by Open Calais. This ID points to the corresponding extracted entity and related tags within this document. Note that this attribute is not relevant to the JSON output format.
Related Tag	em/e/Region

6.5.10 TopmostPublicParentCompany

er/TopmostPublicParentCompany	
Definition	This tag is generated when an extracted entity of the type Company is mapped to a company in the Thomson Reuters dataset that also has a publicly traded parent company. Note that the *openpermid* attribute gives you direct access to high quality, curated Thomson Reuters company data. (In JSON output, the attribute name is *id*.) This metadata type is available upon subscription to Thomson Reuters Intelligent Tagging (TRIT). For more information please contact us at questions@opencalais.com.
Attributes	**commonname**: The commonly used name, as it appears in the Thomson Reuters dataset. **docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format. **id**: See the openpermid attribute description, below. **ispublic**: Indicates whether the company is public (true) or private (false). **legacyid**: The link to the company as it would have been displayed in the legacy Open Calais output. This attribute is relevant to users who upgraded from the legacy Open Calais to the new Open Calais. The purpose of this attribute is to enable you to map the old links, which are no longer supported, to the corresponding new links. This attribute is not relevant to the JSON output format. **name**: The official name as it appears in the Thomson Reuters dataset. **openpermid**: A direct link to the relevant company page on the Open PermID website (https://permid.org). This link gives you direct access to high quality, curated Thomson Reuters company data. (In the JSON output, the attribute name is **id**.).) **permid**: The Thomson Reuters unique ID (PermID) for this company. **primaryric**: The company's primary Reuters Instrument code (RIC) as it appears in the Thomson Reuters dataset. **subject**: The hash tag generated by Open Calais and assigned to the extracted child company. This ID points to the relevant em/e/company and related tags within the document. If the parent company has multiple extracted child companies within the document, additional instances of the subject attribute appear in the tag. Note that this entity is not relevant to the JSON output format. **ticker**: The company stock symbol.
Related Tag	em/e/Company

6.6 Industry Tag

This functionality is available upon subscription to Thomson Reuters Intelligent Tagging (TRIT). For more information please contact us at questions@opencalais.com.

For a conceptual explanation, see Industry Tags.

Industry	
Definition	Identifies the industries related to the companies mentioned in the document.
Attributes	**docid**: The unique ID of the containing document, generated by Open Calais and formatted as a URI. Note that this attribute is not relevant to the JSON output format.
	forenduserdisplay: A recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false).
	name: The name of the industry.
	permid: The Thomson Reuters unique ID (PermID) for this Industry.
	rcscode: The industry code, as defined in the Thomson Reuters Coding Schema (TRCS).
	relevance: Indicates how relevant the industry is to the story. Values range from 0 to 1. The higher the score, the higher the relevance.
	trbccode: The industry code, as defined in the Thomson Reuters Business Classification (TRBC) taxonomy.

Industry

Example

Open Calais Output	*RDF...*

```
<rdf:Description rdf:about="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
    bf3520cddfa6/Industry/5">
    <rdf:type rdf:resource="http://s.opencalais.com/1/type/tag/Industry"/>
    <c:docId rdf:resource="http://d.opencalais.com/dochash-1/7586b818-40af-3d55-ac16-
        bf3520cddfa6"/>
    <c:forenduserdisplay>false</c:forenduserdisplay>
    <c:name>Automobiles &amp; Multi Utility Vehicles</c:name>
    <c:rcscode>B:1294</c:rcscode>
    <c:permid>4294951707</c:permid>
    <c:relevance>0.500</c:relevance>
</rdf:Description>
```

N3...

```
<http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Industry/5>
        a                       <http://s.opencalais.com/1/type/tag/Industry> ;
        c:docId                 <http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-
        816cb722a630> ;
        c:forenduserdisplay    "false" ;
        c:name                  "Automobiles & Multi Utility Vehicles" ;
        c:permid                "4294951707" ;
        c:rcscode               "B:1294" ;
        c:relevance             "0.500" .
```

JSON...

```
"http://d.opencalais.com/dochash-1/268b4e49-098d-389d-9d76-816cb722a630/Industry/5": {
        "_typeGroup": "industry",
        "forenduserdisplay": "false",
        "name": "Automobiles & Multi Utility Vehicles",
        "rcscode": "B:1294",
        "permid": "4294951707",
        "relevance": 0.5
    },
```

Appendix A   Supported Classification Topics

During processing, Open Calais identifies the topics discussed in the document, and generates the relevant DocCat (topic) tags. Currently, the list of possible topics is defined by the <u>Thomson Reuters Coding Schema (TRCS)</u> and/or by the <u>International Press Telecommunications Council (IPTC)</u> news taxonomy.

A.1   Thomson Reuters Coding Schema (TRCS Topics)

Note that access to TRCS topics is available upon subscription to Thomson Reuters Intelligent Tagging (TRIT). For more information please contact us at <u>questions@opencalais.com</u>.

Open Calais supports the following RCS topics:

Note: The **forenduserdisplay** attribute of the DocCat (topic) tag (relevant to allMetadata users only) is a recommendation of whether the tag is suitable as a search item for a specific document (true) or whether the metadata is primarily of use for aggregation and analytics on large quantities of documents (false). Generally speaking, the "true" status is assigned to topics that consistently provide high precision and recall results.

The following list of topics indicates the forenduserdisplay=True topics. All the others are forenduserdisplay=False.

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
1	A:1	Equities Markets	
2	A:2X	Exchange-Traded Funds	
3	A:5	Funds	
4	A6	Real Estate Markets	
5	A:8	Money Markets	
6	A:85	General Obligation Debt	
7	A:86	Revenue Bonds	
8	A:88	Taxables	
9	A:9	Currencies / Foreign Exchange Markets	
10	A:A	Asset-Backed Securities	
11	A:D	Mortgage-Backed Markets	
12	A:E	Loans	
13	A:F	Hedge Funds	True
14	A:G0	Sovereign Wealth Funds	
15	A:N	National Government Debt	
16	A:P	Equity-Linked Bonds	
17	A:Q	Private Equity Funds	
18	A:S	Eurobonds	
19	A:T	Corporate Debt	

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
20	A:U	Credit and Corporate Debt	
21	A:W	Mortgage Agency Debt	
22	A:X	High-Yield Securities	
23	A:Y	Investment Grade Securities	
24	B:108	Beverages (TRBC)	
25	B:109	Brewers (TRBC)	
26	B:110	Distillers & Wineries (TRBC)	
27	B:113	Fishing & Farming (TRBC)	
28	B:114	Food Processing (TRBC)	
29	B:115	Tobacco (TRBC)	
30	B:125	Financials (TRBC)	
31	B:126	Banking & Investment Services (TRBC)	
32	B:128	Banks (TRBC)	
33	B:129	Consumer Financial Services (TRBC)	
34	B:130	Investment Banking & Investment Services (TRBC)	
35	B:136	Insurance (TRBC)	
36	B:14	Chemicals (TRBC)	
37	B:144	Residential & Commercial REITs (TRBC)	
38	B:148	Healthcare (TRBC)	
39	B:156	Pharmaceuticals & Medical Research (TRBC)	
40	B:16	Agricultural Chemicals (TRBC)	
41	B:161	Technology (TRBC)	
42	B:162	Technology Equipment (TRBC)	
43	B:163	Semiconductors & Semiconductor Equipment (TRBC)	
44	B:164	Semiconductors (TRBC)	
45	B:167	Communications Equipment (TRBC)	
46	B:168	Computers & Office Equipment (TRBC)	
47	B:172	Software & IT Services (TRBC)	
48	B:174	Software (TRBC)	
49	B:177	Telecommunications Services (TRBC)	
50	B:181	Utilities (TRBC)	
51	B:185	Natural Gas Utilities (TRBC)	
52	B:187	Water & Other Utilities (TRBC)	

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
53	B:19	Mineral Resources (TRBC)	
54	B:2	Energy (TRBC)	
55	B:20	Metals & Mining (TRBC)	
56	B:208	Investment Management & Fund Operators (TRBC)	
57	B:21	Precious Metals & Minerals (TRBC)	
58	B:211	Financial & Commodity Market Operators (TRBC)	
59	B:22	Steel (TRBC)	
60	B:23	Aluminum (TRBC)	
61	B:26	Construction Materials (TRBC)	
62	B:28	Paper & Forest Products (TRBC)	
63	B:29	Forest & Wood Products (TRBC)	
64	B:37	Aerospace & Defense (TRBC)	
65	B:38	Machinery, Equipment & Components (TRBC)	
66	B:4	Coal (TRBC)	
67	B:5	Oil & Gas (TRBC)	
68	B:56	Transportation (TRBC)	
69	B:58	Air Freight & Courier Services (TRBC)	
70	B:59	Airline Services (TRBC)	
71	B:60	Airlines (TRBC)	
72	B:62	Marine Services (TRBC)	
73	B:63	Freight Transportation, Marine (TRBC)	
74	B:64	Marine Port Services (TRBC)	
75	B:65	Transportation, Ground (TRBC)	
76	B:66	Passenger Transportation, Ground & Sea (TRBC)	
77	B:67	Freight Transportation, Ground (TRBC)	
78	B:68	Highways & Rail Tracks (TRBC)	
79	B:7	Oil & Gas Exploration and Production (TRBC)	
80	B:71	Automobiles & Auto Parts (TRBC)	
81	B:72	Auto & Truck Manufacturers (TRBC)	
82	B:73	Auto, Truck & Motorcycle Parts (TRBC)	
83	B:74	Tires & Rubber Products (TRBC)	
84	B:78	Consumer Electronics (TRBC)	
85	B:82	Textiles & Apparel (TRBC)	

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
86	B:83	Textiles & Leather Goods (TRBC)	
87	B:84	Apparel & Accessories (TRBC)	
88	B:85	Footwear (TRBC)	
89	B:87	Hotels & Entertainment Services (TRBC)	
90	B:90	Casinos & Gaming (TRBC)	
91	B:91	Leisure & Recreation (TRBC)	
92	B:92	Media & Publishing (TRBC)	
93	B:93	Advertising & Marketing (TRBC)	
94	B:94	Broadcasting (TRBC)	
95	B:95	Entertainment Production (TRBC)	
96	B:96	Publishing (TRBC)	
97	B:98	Retailers (TRBC)	
98	E:1K	Share Consolidations	True
99	E:1O	Indices Changes	True
100	E:2P	Earnings Releases	
101	E:2X	Shareholder Activism	True
102	E:3Y	Labour Disputes	True
103	E:3Z	Job Cuts	True
104	E:44	Commercial Contract Wins / Terminations	True
105	E:45	Monopolies / Antitrust Issues	True
106	E:4B	Key Personnel Changes	
107	E:4C	Name Changes	True
108	E:4D	Shareholder Meetings	True
109	E:4E	Accounting Issues	True
110	E:4G	Strategic Combinations	
111	E:4J	Divestitures / Spin-Offs	True
112	E:4L	Placings	True
113	E:4M	Privatisations	
114	E:4N	Share Splits	True
115	E:4P	Share Buybacks/Repurchases	True
116	E:4R	Director Dealings	True
117	E:4S	Currency Intervention	
118	E:4T	Currency Revaluation	

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
119	E:4V	Corporate Debt Restructurings	True
120	E:5	Economic Events	
121	E:56	Organizational Restructuring	True
122	E:57	Corporate Litigation	True
123	E:58	Earnings Calls/Guidance Calls	
124	E:5A	Equity Stakes	True
125	E:5B	Bonus Shares Issues	True
126	E:5C	Secondary Share Offerings / Issues	True
127	E:5D	Expansions/New Markets/New Units	True
128	E:5G	Security Listings / Delistings	True
129	E:5I	Bond Sales	
130	E:5J	Production Status Changes	
131	E:6	Market Events	
132	E:61	Block Trades	True
133	E:64	Banking Capital and Liquidity Requirements	True
134	E:66	Product Recalls	True
135	E:67	Same-Store Sales	True
136	E:68	Change of CEO	True
137	E:69	Change of CFO	True
138	E:6A	Change of Chairperson	True
139	E:6B	Change of President	True
140	E:6N	Offers For Sale	True
141	E:7	Class Actions	True
142	E:A	Interest Rates / Policy	
143	E:B	Monetary / Fiscal Policy / Policy Makers	
144	E:C	Labour / Personnel	
145	E:D	Corporate/ Market Regulation	True
146	E:E	Management Issues / Policies	
147	E:F	Dividends	
148	E:G	Performance / Results / Earnings	True
149	E:H	Bankruptcy/Insolvency	True
150	E:J	Mergers / Acquisitions / Takeovers	True
151	E:K	Equity Financing	

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
152	E:L	Broker Research / Recommendations	True
153	E:M	Initial Public Offering	
154	E:Q	Results Forecasts / Warnings	True
155	E:R	Deals	
156	E:S	Competitive Bond Sales	
157	E:V	Negotiated Sales	
158	E:W	New Issues	True
159	F:2	Pipelines	
160	F:49	Oil Shale	
161	F:C	Transmission Lines	
162	G:19	Denmark	
163	G:1C	Slovakia	
164	G:1G	Jamaica	
165	G:1J	Sri Lanka	
166	G:1K	Scotland	
167	G:1L	Cook Islands	
168	G:1R	Wales	
169	G:1X	Montserrat	
170	G:1Y	Solomon Islands	
171	G:21	England	
172	G:22	Haiti	
173	G:26	Brazil	
174	G:28	Yemen	
175	G:2C	Tuvalu	
176	G:2E	Czech Republic	
177	G:2G	Papua New Guinea	
178	G:2I	South Africa	
179	G:2J	Nepal	
180	G:2K	Venezuela	
181	G:2M	Lesotho	
182	G:2N	Palau	
183	G:2P	Pakistan	
184	G:2R	Greenland	

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
185	G:2S	Colombia	
186	G:2V	Mexico	
187	G:2X	Mauritania	
188	G:2Z	Romania	
189	G:3	Western Europe	
190	G:30	Switzerland	
191	G:31	Antigua and Barbuda	
192	G:36	Marshall Islands	
193	G:39	American Samoa	
194	G:3A	Saint Lucia	
195	G:3D	Germany	
196	G:3E	Montenegro	
197	G:3F	Saint Vincent and the Grenadines	
198	G:3H	Hong Kong	
199	G:3I	Macau	
200	G:3N	Norway	
201	G:3S	Israel	
202	G:3T	Peru	
203	G:3U	Liberia	
204	G:3Z	Fiji	
205	G:4	Americas	
206	G:41	Japan	
207	G:44	Guyana	
208	G:49	New Zealand	
209	G:4F	Western Sahara	
210	G:4G	Malta	
211	G:4H	Latvia	
212	G:4I	Cameroon	
213	G:4J	Cayman Islands	
214	G:4L	Wallis and Futuna	
215	G:4N	Tajikistan	
216	G:4P	Syria	
217	G:4Q	Guadeloupe	

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
218	G:4R	Azerbaijan	
219	G:4Y	Uruguay	
220	G:52	Zimbabwe	
221	G:54	French Polynesia	
222	G:55	Spain	
223	G:59	Palestinian Territories	
224	G:5B	India	
225	G:5C	Seychelles	
226	G:5G	Central African Republic	
227	G:5H	Costa Rica	
228	G:5K	Congo-Brazzaville / Republic of Congo	
229	G:5L	Equatorial Guinea	
230	G:5M	France	
231	G:5N	Ghana	
232	G:5X	Croatia	
233	G:5Y	Poland	
234	G:60	Argentina	
235	G:62	Niue	
236	G:68	Burundi	
237	G:6B	Nigeria	
238	G:6C	Belize	
239	G:6E	Senegal	
240	G:6G	Malawi	
241	G:6I	Iceland	
242	G:6K	Eritrea	
243	G:6L	Ethiopia	
244	G:6Q	Namibia	
245	G:6V	Sweden	
246	G:6W	Libya	
247	G:6Z	Djibouti	
248	G:72	Myanmar	
249	G:74	Slovenia	
250	G:75	Bermuda	

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
251	G:77	Gambia	
252	G:7B	Oman	
253	G:7E	British Virgin Islands	
254	G:7G	Albania	
255	G:7H	Swaziland	
256	G:7J	United Kingdom	
257	G:7K	Netherlands	
258	G:7M	Luxembourg	
259	G:7N	Georgia	
260	G:7P	Kiribati	
261	G:7R	Afghanistan	
262	G:7S	Algeria	
263	G:7U	Taiwan	
264	G:7X	Cuba	
265	G:7Y	Norfolk Island	
266	G:83	North Korea	
267	G:85	Kazakhstan	
268	G:86	Suriname	
269	G:88	Monaco	
270	G:89	Paraguay	
271	G:8A	Congo-Kinshasa / Democratic Republic of Congo	
272	G:8C	Martinique	
273	G:8J	Nauru	
274	G:8K	Tonga	
275	G:8P	Barbados	
276	G:8R	Kyrgyzstan	
277	G:8T	Cyprus	
278	G:8U	Dominica	
279	G:8W	Canada	
280	G:8X	Morocco	
281	G:90	Finland	
282	G:91	Togo	
283	G:92	Saudi Arabia	

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
284	G:96	Guatemala	
285	G:9A	Grenada	
286	G:9D	Estonia	
287	G:9E	Micronesia	
288	G:9G	Cambodia	
289	G:9M	Vanuatu	
290	G:9R	British Indian Ocean Territory	
291	G:9T	Trinidad and Tobago	
292	G:9U	Kosovo	
293	G:9W	United States Minor Outlying Islands	
294	G:9Y	Belgium	
295	G:A	Europe	
296	G:A3	Portugal	
297	G:A4	United Arab Emirates	
298	G:A6	Chad	
299	G:A7	Northern Ireland	
300	G:AC	Nicaragua	
301	G:AD	Aruba	
302	G:AE	South Korea	
303	G:AI	Macedonia	
304	G:AY	Antarctica	
305	G:B	Central / Eastern Europe	
306	G:B2	East European Countries	
307	G:B4	Eurozone as a Whole	
308	G:M	Nordic States	
309	G:U	Baltic States	
310	I:11	Money Supply	
311	I:18	Consumer Sentiment	
312	I:19	Industrial Output	
313	I:1A	Business Sentiment	
314	I:1C	Home Sales	
315	I:1D	Mortgage Applications	
316	I:1E	Bankruptcy Figures	

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
317	I:1N	Consumer Credit	
318	I:1P	Employment / Unemployment	
319	I:1Q	Government Borrowing Requirement	
320	I:1R	Inventories	
321	I:1U	Trade / Current Account	
322	I:1W	Capital Movements	
323	I:1Y	Official Reserves	
324	I:B	Inflation	
325	M:18	Retirement / Old Age	
326	M:1LY	Disasters/Accidents	True
327	M:1M3	Medical Regulatory Issues	True
328	M:2	Military Conflicts	
329	M:24	Tourism / Travel	
330	M:3	Arts / Culture / Entertainment	
331	M:4	Asylum / Immigration / Refugees	
332	M:4Y	Cross-Country Skiing	
333	M:50	Champions League	
334	M:5C	Archery	
335	M:5D	Athletics	
336	M:5E	Australian Rules Football	
337	M:5G	Baseball	
338	M:5H	Basketball	
339	M:5J	Biathlon	
340	M:5K	Billiards	
341	M:5Q	Canoeing	
342	M:5V	Cycling	
343	M:60	Fencing	
344	M:65	Handball	
345	M:66	Field Hockey	
346	M:6B	Karate	
347	M:6G	Modern Pentathlon	
348	M:6H	Motor Racing	
349	M:6J	Nordic Skiing	

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
350	M:6S	Rugby Union	
351	M:6T	Shooting	
352	M:6Y	Soccer	
353	M:70	Speed Skating	
354	M:72	Squash	
355	M:75	Swimming	
356	M:76	Table Tennis	
357	M:78	Tennis	
358	M:79	Tenpin Bowling	
359	M:7H	Yachting / Sailing	
360	M:7I	NBA Basketball	
361	M:7J	NFL American Football	
362	M:7S	Association of Southeast Asian Nations	
363	M:8	Central Banks / Central Bank Events	
364	M:8E	North Atlantic Treaty Organization	
365	M:9	Civil Unrest	
366	M:9N	World Bank	
367	M:9R	United States Department of Agriculture	
368	M:A	Crime	
369	M:AJ	US Government News	
370	M:B	Defense	
371	M:B6	Mining	
372	M:C	Diplomacy / Foreign Policy	
373	M:E9	Government Finances	
374	M:EG	Tariffs	
375	M:EI	Subsidies	
376	M:EL	Conflicts / War / Peace	
377	M:EM	Overseas Development Aid	
378	M:EV	Judicial Process / Court Cases / Court Decisions	
379	M:F	Education	
380	M:F2	Kidnapping	
381	M:F3	Sex Crimes	
382	M:F4	Murder / Manslaughter	

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
383	M:F7	Bombing (Non-Military)	
384	M:FD	Computer Crime / Hacking / Cybercrime	True
385	M:FF	Corruption / Bribery / Embezzlement	True
386	M:FJ	Drug Trafficking / Narcotics	
387	M:FX	Workers Pay	True
388	M:G	Elections / Voting	
389	M:GM	Management Pay	True
390	M:GW	Credit Enhancement	
391	M:GX	Terms / Conditions	
392	M:GZ	Art	
393	M:H	Environment	
394	M:H0	Music	
395	M:H1	Film	
396	M:H8	Fashion	
397	M:I9	Religion / Belief	
398	M:IA	Christianity	
399	M:IG	Islam	
400	M:IZ	Lesbian / Gay / Bisexual / Transsexual	
401	M:J2	Women's Issues	
402	M:JZ	Pollution	True
403	M:K	European Union	
404	M:K2	Genetically Modified Organisms	
405	M:KF	Fires	
406	M:KJ	Floods	
407	M:KL	Earthquakes	
408	M:KM	Wildfires / Forest Fires	
409	M:KN	Droughts	
410	M:L6	Olympics-Factboxes	
411	M:L8	Olympics-Medals	
412	M:LB	Olympics-Results	
413	M:LD	World Cup	
414	M:LK	International / National Security	
415	M:M0	Human Rights / Civil Rights	True

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
416	M:MT	Sports Business	
417	M:MU	International Criminal Court	
418	M:MW	Non-Governmental Organizations	
419	M:MX	Aid Relief / Humanitarian Agencies	
420	M:MY	Advocacy Groups / Pressure Groups / Lobbies	
421	M:MZ	Nuclear Armaments / Nuclear Proliferation	
422	M:N	Government / Politics	
423	M:N1	Anti-competitive Behaviour / Price Fixing	True
424	M:N2	Financial Fraud / Securities Fraud	
425	M:N3	Intellectual Property Crime	True
426	M:N4	Climate Politics	
427	M:N5	Fundamental Rights / Civil Liberties	
428	M:N7	Gun Control	
429	M:N8	Privacy / Data Protection	True
430	M:N9	Military Procurement	
431	M:NA	Air Accidents	
432	M:NC	Ground Accidents / Collisions	
433	M:ND	Sea Accidents	
434	M:NI	Nature / Wildlife	
435	M:NK	Tsunami	
436	M:NL	Information Technologies / Computer Sciences	
437	M:NM	Life Sciences	
438	M:NW	Children / Youth Issues	
439	M:NX	Insurgencies	
440	M:P	Health / Medicine	
441	M:R	International Trade	True
442	M:S	Internet / World Wide Web	
443	M:T	Lawmaking	
444	M:U	Olympics	
445	M:V	Science	
446	M:X	Taxation	True
447	U:13	Rubber	
448	U:19	Cobalt	

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
449	U:1A	Ruthenium	
450	U:1D	Silicon	
451	U:1F	Vanadium	
452	U:1K	Lead	
453	U:1N	Peanut Oil	
454	U:1Z	Rye	
455	U:22	Coconut Oil	
456	U:23	Gasoline	
457	U:24	Refined Products	
458	U:29	Cadmium	
459	U:2A	Indium	
460	U:2B	Barge	
461	U:2H	Stainless Steel	
462	U:2I	Chromium	
463	U:2J	Tin	
464	U:2L	Olive Oil	
465	U:2R	Natural Gas	
466	U:2Z	Liquefied Petroleum Gas	
467	U:3	Base Metals	
468	U:30	Soybean	
469	U:3H	Sugar	
470	U:3K	Corn Oil	
471	U:3N	Magnesium	
472	U:3T	Gallium	
473	U:3V	Orange Juice	
474	U:3Y	Bismuth	
475	U:3Z	Rice	
476	U:40	Soybean Oil	
477	U:41	Milk	
478	U:42	Molybdenum	
479	U:43	Selenium	
480	U:44	Tungsten	
481	U:45	Crude Oil	

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
482	U:47	Palm Oil	
483	U:4A	Potash	
484	U:4C	Dry Bulk Freight	
485	U:4G	Germanium	
486	U:4H	Sorghum	
487	U:4T	Rapeseed Oil	
488	U:4U	Cotton	
489	U:51	Gold	
490	U:52	Cheese	
491	U:55	Iridium	
492	U:5A	Palladium	
493	U:5B	Wool	
494	U:5C	Wine	
495	U:5E	Manganese	
496	U:5J	Sunflower Oil	
497	U:5K	Tanker Freight	
498	U:5M	Bunker Fuel	
499	U:5U	Jet Fuel / Kerosene	
500	U:5V	Wheat	
501	U:6B	North Sea Crudes	
502	U:6D	Barley	
503	U:6F	Corn	
504	U:6G	Plastics / Polymers	
505	U:6K	Titanium	
506	U:6L	Heating Oil / Gas Oil	
507	U:6Q	Zinc	
508	U:6U	Coffee	
509	U:6V	Platinum	
510	U:6Z	Cocoa	
511	U:73	Antimony	
512	U:76	Silk	
513	U:77	Petrochemicals	
514	U:7F	Uranium	

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
515	U:7G	Biodiesel	
516	U:7H	Cellulosic Biofuels	
517	U:7I	Bioethanol	
518	U:8C	Arabica	
519	U:8K	Latin America Crudes	
520	U:9K	South East Asia Crudes	
521	U:9T	Europe / Middle East / Africa Crudes	
522	U:9U	Africa Crudes	
523	U:B	Transportation Markets / Transportation	
524	U:B5	Russia / Other CIS Crudes	
525	U:BN	Hydroelectric Power Stations	
526	U:BP	Solar Power Stations	
527	U:BS	Wind Farms	
528	U:BV	Iron Ore	
529	U:C	Energy Markets	
530	U:C2	Scrap Metals	
531	U:CA	Diamond Markets	
532	U:CM	Container	
533	U:CN	Road Freight	
534	U:CX	Precipitation	
535	U:D	Livestock	
536	U:D0	Temperature	
537	U:D2	Wind / Hurricanes / Typhoons / Tornadoes	
538	U:DR	North Asia Crudes	
539	U:E	Emissions Markets	
540	U:EP	Phosphates	
541	U:ES	Coal-fired Power Stations	
542	U:ET	Gas-fired Power Stations	
543	U:EU	Oil-fired Power Stations	
544	U:F	Oilseeds	
545	U:F1	Water	
546	U:F2	Spices / Herbs / Fruit / Vegetables	
547	U:FA	Steel	

	RCS CODE	TOPIC NAME	FORENDUSERDISPLAY
548	U:FF	Platinum Group Metals	
549	U:H	Coal	
550	U:I	Weather Markets / Weather	
551	U:K	Meals / Feeds / Pulses	
552	U:T	Iron / Steel	

A.2   IPTC Topics (News)

Open Calais supports a small set of high-level news topics defined by the International Press Telecommunications Council (IPTC) news taxonomy. The IPTC topics currently supported by Open Calais are:

- **Business_Finance**: corporate financial results, joint business ventures, global currencies, prices and markets, stocks and bonds, prices, economic forums.

- **Disaster_Accident**: man-made and natural events resulting in damage to objects, loss of life or injury.

- **Education**: topics related to aspects of furthering knowledge of humans.

- **Entertainment_Culture**: media, movies and TV, literature and journalism, music, celebrities, entertainment products, internet culture, youth culture.

- **Environment**: topics related to the condition of our planet such as natural disasters, protection, and their effect on living species as well as inanimate objects or property.

- **Health_Medical_Pharma**: hospitals and healthcare, medical research, diseases, drugs, pharmaceutical industry, health insurance, diet and nutrition.

- **Hospitality_Recreation**: eating and travel, leisure/recreational facilities and general activities undertaken for pleasure and relaxation.

- **Human Interest**: lighter topics of general interest for humans.

- **Labor**: topics related to the employment of individuals, support of the unemployed.

- **Law_Crime**: topics relating to the enforcement of rules of behavior in society, breaches of these rules and the resulting punishments; law firms, legal practice and lawsuits.

- **Politics**: government policies and actions, politicians and political parties, elections, war and acts of aggression between countries.

- **Religion_Belief**: theology, philosophy, ethics and spirituality.

- **Social Issues**: topics related to aspects of the behavior of humans affecting the quality of life.

- **Sports**: sports competitions and tournaments, athletes, Olympic games.

- **Technology_Internet**: technological innovations, technology-related companies, hardware and software products, internet products and web sites, telecom industry.

- **Weather**: topics relating to meteorological phenomena.

- **War_Conflict**: topics related to acts of socially- or politically- motivated protest and/or violence.

- **Other**: miscellaneous topics not covered by any of the other categories.

Appendix B   Third Party NLP Onboarding

The Third Party Onboarding mechanism is available to users of Open Calais.

Open Calais supports integration with third party NLP services. Open Calais is easily configured so that API calls it receives are processed both by Open Calais and by the third party NLP service in parallel; the results of both requests are concatenated and returned to the user in a single Open Calais RDF output file.

The file is structured as follows:

```
<rdf:RDF>
        Open Calais response

      <!--*****  THIRD PARTY RESULTS ***** -->

        Third party response

      </rdf:RDF>
```

As illustrated above, the RDF file is comprised of two parts: the Open Calais output, followed by the third party NLP output.

For a detailed explanation of the Open Calais response, see RDF Response Format. See also The API Response.

The third party NLP output has no relation to the Open Calais output or schema.

If either of the two parallel requests cannot be processed due to format or load issues, an error message is returned in the relevant part of the RDF file. An error generated by one of the processing engines does not affect the response of the other processing engine. In other words, one part of the RDF output file may be populated by metadata, while the other part may display an error message.

THOMSON REUTERS