

RavenPack Analytics

User Guide and Service Overview

Powered by



RavenPack

CONFIDENTIAL INFORMATION

Version 1.0

Last updated: May 18, 2020

TABLE OF CONTENTS

Introduction	2
Key Benefits	2
How and Where Are Analytics Produced?	3
Where to Find Analytics	3
Entity, Event and Topic Detection	4
Automated Content Analysis	6
Latency	6
Key Information	7
Data File Contents	7
Data Field Descriptions	8
Functions (Indicators)	20
Document Search Filters	24
Alerts	24
RavenPack Reference Service	25
The RavenPack Taxonomy	29
Classification Structure	30
Properties	30
Fact Level	30
Categories	31
The Benefits	31
The Event Taxonomy File	31
Dataset Schemas	33

Introduction

RavenPack Analytics transforms large unstructured datasets, such as traditional news and social media, into structured granular data and indicators to help financial services firms improve their performance. The product serves to overcome the challenges posed by the characteristics of Big Data - volume, variety, veracity, and velocity - by converting unstructured content into a format that can be more effectively analyzed, manipulated, and deployed in financial applications. Whether your objective is generating more alpha, managing risk more effectively, cutting false positives in market surveillance or generating trading ideas, RavenPack Analytics can improve your performance.

Financial firms can use RavenPack data to lead portfolio allocations, which help improve the average life and profit of existing trading models. They can use it to enhance risk-adjusted returns from trading or investing, to manage event risk when investing or market-making, to assist compliance and surveillance analysts, and to encourage trading activity. These analytics not only allow market participants to capture alpha opportunities, but they also help to improve risk management and provide for better trading execution.

Key Benefits

RavenPack Analytics can protect portfolio managers or traders from the consequences of missing important news that has an impact on their position or portfolio. Also, news events on natural disasters, economic indicators, earnings, product recalls, layoffs, stock or credit ratings, and many others can be precursors to changes in volatility of securities. RavenPack Analytics enables traders to get an edge by acting in advance of these changes.

Traders also boost their gains using news-based algorithms to speed their response time to breaking events. They build defensive applications to ensure that key news events are factored in and use RavenPack's automated news analysis as a more effective form of low latency decision support. Firms reduce the time required for low-frequency fundamental traders to assess their options manually and execute their responses more effectively. RavenPack Analytics can also help in post-trade analysis to explain why a traditional algorithm or strategy did not work.

With RavenPack, it is now possible to measure and instantly incorporate business, economic, and geopolitical events that are difficult to predict and often destabilizing. Wars, elections, terrorist attacks, airspace closures, or volcanic ash clouds – RavenPack monitors every event, all the time, so you're never caught off guard.

How and Where Are Analytics Produced?

People have been "manually" extracting patterns from news for many years, but the increasing volume of news in modern times has called for more automated approaches. In RavenPack Analytics, news is received and processed through an automated real-time data stream made up of several software components, some running in parallel. Once content analysis and sentiment calculations have been made, information about the news is immediately sent on to subscribers as real-time events, which they use to drive trading models and applications.

RavenPack processes content in the Amazon cloud. Amazon manages all data centers, servers, networks, and storage systems while RavenPack focuses on building software that runs in this environment. The Amazon Cloud allows many servers to be run in parallel in order to handle the large amount of distributed processing that must take place in order to produce RavenPack Analytics.

The following sections describe the process involved in producing news analytics including details on RavenPack's proprietary sentiment analysis techniques.

Where to Find Analytics

RavenPack Analytics may be accessed using a variety of different methods:

RavenPack Platform

RavenPack has a platform that allows analytics data to be organized as datasets and either requested as flat-files, or visualized on dashboards as time series, treemaps and other charts. You can access the RavenPack platform at:

<https://ravenpack.com/>

RavenPack Web API

The RavenPack Web API may be used for programmatically managing datasets and directly requesting data dumps or subscribing to real-time data using our REST API. Ask your RavenPack sales representative for more information.

Full Historical Archives

Archives are available that contain all RavenPack analytics granular data from 2000 to present. Ask your RavenPack sales representative for more information.

We set limits and quotas on requests for RavenPack Analytics in order to ensure a fair distribution of system resources and to protect the system from receiving more requests than it is able to handle. These limits and quotas are subject to change.

The following quotas apply to requests made via the RavenPack Platform, Web API and Streaming API:

- 10 simultaneous subscriptions per user to the streaming API
- 10 simultaneous datafile requests per user

There are also additional limits in place for daily data allowances based on a rolling 24-hour window. You will be notified with the following message via the Platform or API should you exceed this allowance:

“Your datafile requests have exceeded the daily data allowance over the past 24 hours. If you believe you are receiving this message in error please contact support@ravenpack.com”

Entity, Event and Topic Detection

When dealing with the tens of thousands of stories published about entities every day, it makes sense to try to classify them into a set of predefined categories. RavenPack has applied a number of proprietary technological solutions that categorizes stories into a simple set of themes which are fundamental to today’s investment environment.

The technique of producing this kind of analysis came about by performing a careful study of the types of stories available on entities and by extracting the primary categories that would allow meaningful interpretation of a story. Once the categories had been determined, the goal was to implement technology that could perform the classification automatically. Some categories are more straightforward than others, thus different techniques were applied.

Events in RavenPack’s Taxonomy are defined using thousands of proprietary template programs and Part-of-Speech tagging. Templates are compositions of language tokens or values taken in a specific context. Tokens may be a type of language marker, such as a number or date. They may be words or phrases, perhaps broken down to their root form or taken only for a given tense.

Part-of-speech tagging involves marking up the words in a text corpus as corresponding to a particular part of speech, based on both its definition, as well as its context - i.e. relationship with adjacent and related words in a phrase, sentence, or paragraph. This makes templates more scalable, modular, and effective.

For example, one template for the event category *bankruptcy* is defined as:

Citadel Broadcasting Files for Bankruptcy in New York
(`$COMPANY %FILE FOR %BANKRUPTCY IN $PLACE`)

Likewise, a template for the event category *sec-investigation* might be defined as:

Xerox Sues Google, Yahoo over Search-Query Patents –Report
(`$ENTITY %SUE $ENTITY $ENTITY %PREPOSITION %PATENT-TYPE`)

The classification is broken up into different steps:

1. **Identification of entities mentioned in the story:**

RavenPack's proprietary technology is applied to the full document text to identify and match entities to corresponding identifiers within a point-in-time sensitive database. This process takes into account any available metadata provided in the document, such as inferred or related entities and other known aliases.

2. **Extraction of the event type:**

The story is broken up into its linguistic components, e.g. words, phrases, parts-of-speech, etc. Further attribution is carried out to apply semantic meaning and a set of rules are applied to determine the different types of events taking place across the text of the story.

3. **Role Detection:**

Next, the role that each entity plays within each event is deduced from the context in which the entities are mentioned. It is possible for multiple entities to be assigned roles within an event, for example, the defendant and the plaintiff within a legal issues event.

4. **Consolidation of Information:**

Consolidation is the last step in the process, allowing multiple detections or inferences of entities and events, within the same document, to be combined and assigned as the final event category for the entity/ies.

A full list of categories and definitions that form part of the RavenPack's Taxonomy is available in the Product Area.

Additional information regarding RavenPack's Event Detection Technology can be found in this [Technical Note](#).

Automated Content Analysis

Many of the RavenPack's classifiers are effectively source-independent and will work without problems on a textual story retrieved from any publisher service or website. Others are trained or configured for one particular source and need to be customized in order to get the best results. Some classifiers are designed on particular topics or themes and these will target specific sources only. Classifiers are RavenPack's proprietary natural language processing software deployed across multiple high-speed servers in virtual machines.

The result of automated content analysis is an aggregated data feed comprised entirely of structured news information, enriched with RavenPack's metadata including tags on entities, events, entity roles, novelty, sentiment, and other analytics - ready for consumption by all types of applications.

Latency

Text Analysis Latency

The Latency of the RavenPack text analysis process is defined as the time it takes an average story to pass through the classification system, from the time RavenPack receives it, to the time at which it has been fully classified. RavenPack has calculated the expected time to process a document in isolation as 300ms.

The steps involved to process a story are as follows:

- 1: Document arrives in the collector and is normalized into a common format
- 2: Document arrives in the classifiers and is processed using multiple techniques
- 3: Document arrives in the Data Feed Server with complete analytics

Real-time Throughput

RavenPack processes millions of documents each day, simultaneously classifying tens to thousands of documents. The overall latency of the real-time feed may vary depending on the content load that the system is processing at any one time, which can be affected by larger documents and high volume bursts of content being published.

Under normal load conditions the real-time latency that can be expected from the real-time feed can range from 250 to 600ms.

Delivery Latency

The additional transmission latency between RavenPack's server and subscribed clients are outside of RavenPack's control and depend on a number of different factors such as geographical distance, network quality, network speed etc. It is possible to determine the transmission latency using tools like ping.

Key Information

RavenPack Analytics delivers sentiment analysis and event data most likely to impact financial markets and trading around the world. The service includes analytics on more than 300,000 entities in over 130 countries and covers over 98% of the investable global market. All relevant news items about entities are classified and quantified according to their sentiment, relevance, topic, novelty, and market impact; the result is a data product that can be segmented into many distinct benchmarks and used in a variety of applications.

Entity Type Coverage

Places:	138,700+
Companies:	86,000+
People:	11,500+
Organizations:	5,500+
Products:	58,000+
Positions:	500+
Sports Teams:	280+
Nationalities:	230+
Currencies:	1,000+
Commodities:	100+

Equity Entity Coverage by Region

Americas:	49.60%
Asia:	23.00%
Europe:	21.80%
Oceania:	4.90%
Africa:	0.70%

Data File Contents

RavenPack Analytics is comprised of content from premium newswires (such as Dow Jones, Benzinga, MT Newswires, Alliance News, FX Street, The Fly), many providers of regulatory news and press releases, and over 22,000 web publications.

Source:	Combines content from over 22,000 sources
Data Format:	Comma separated values (.csv) files
File Encoding:	UTF-8
Archive Start:	Jan 1, 2000
Archive Packaging:	Monthly .csv files in yearly .zip archives

Data Fields: 50 fields

Whenever an entity such as a company, organization, product, or currency is mentioned in the news, RavenPack produces records about those entities. A single news story can yield multiple records if more than one entity is mentioned.

Note: Please refer to the RavenPack Reference Service section of this guide for the structure and formatting of reference data files.

Data Field Descriptions

TIMESTAMP.UTC

The Date/Time (YYYY-MM-DD hh:mm:ss.sss) at which the news item was received by RavenPack servers in Coordinated Universal Time (UTC).

TIMESTAMP.TZ

The Date/Time (YYYY-MM-DD hh:mm:ss.sss) at which the news item was received by RavenPack servers in the time zone specified by the user.

RP_STORY_ID

A 32 character alphanumeric identifier that uniquely identifies each document in the RavenPack Analytics data. Example: 1FB2B3F5E99C4D3BCF59FDB3E8C8C9BD.

RP_ENTITY_ID

A unique and permanent entity identifier assigned by RavenPack. Every entity tracked is assigned a unique identifier comprised of 6 alphanumeric characters. Example: 228D42.

The RP_ENTITY_ID field consistently identifies entities throughout the historical archive. RavenPack's entity detection algorithms find only references to entities by information that is accurate at the time of story publication (point-in-time sensitive). A full list of RP_ENTITY_IDs is available via the Entity Mapping API.

ENTITY_TYPE

The type of entity associated with a particular RP_ENTITY_ID. Currently RavenPack produces analytics on the following types of entities:

Type	Name	Description
CMDT	Commodity	Exchange traded commodities such as crude oil and soy.

COMP	Company	Business organizations that may be traded on a stock exchange.
CURR	Currency	Currencies of all financial and industrial countries.
NATL	Nationality	The status of belonging to a particular nation.
ORGA	Organization	Non-business organizations including government entities.
PEOP	People	Individuals that are mentioned in the news.
PLCE	Place	Towns, cities, and countries.
PROD	Product	Products and services.
PRDT	Product Type	A categorization of related products.
TEAM	Sports Teams	Professional teams from a variety of different sports.

ENTITY_NAME

The official canonical name of the entity identified by the RP_ENTITY_ID.

COUNTRY_CODE

The two character ISO-3166 country code associated with an entity. Companies and organizations are associated with the country of incorporation, currencies are associated with the country where the central bank resides. Entities such as commodities have a country code of XX as they are global and not associated with any specific country.

RELEVANCE

An integer score between 0-100 that indicates how strongly related the mention of an entity is to the underlying news story, with higher values indicating greater relevance. For any news story that mentions an entity, RavenPack provides a relevance score. A score of 0 means the entity was passively mentioned while a score of 100 means the entity was prominent in the news story. Values above 75 are considered significantly relevant.

RavenPack’s analysis is not limited to keywords or mentions when calculating relevance. Automated classifiers look for meaning by detecting the roles entities play in specific events like acquisitions or legal disputes or when announcing corporate actions, executive changes, product launches, or recalls, among many other event types.

An entity will be assigned a high mark of 100 if a) it is found to be playing a key role in the first event detected in the headline of a story, and b) it is not playing an explicitly lower relevant role such as a rater. If an entity is otherwise referenced in the headline or story body, it will receive a value between 0 and 99. The score is assigned by a proprietary text positioning algorithm based on events in which the entity plays a key role, the paragraph in which the entity is mentioned (i.e. headline, first paragraph, second paragraph, etc.), the number of references in the text, and the

overall number of entities mentioned in the story. Usually, a relevance value of at least 90 indicates that the entity is referenced in the main title or headline of the news item, while lower values indicate references further in the story body.

For example, if the headline reads *“JP Morgan upgrades IBM from Hold to Buy and Microsoft hires new CFO”*, IBM would receive a relevance score of 100 since it plays a key role in the first event in the headline of the story. In this example, JP Morgan would get a score between 90 and 99 because, as a rater, it does not play a key role in the first event. Similarly, Microsoft is detected in an event in the headline of the story, but is not detected playing a key role in the first event, therefore it will also receive a relevance score between 90 and 99. When IBM is mentioned only in the body of the story, it will receive a relevance score anywhere between 0 and 89, depending on how often it was mentioned and in which paragraphs it was mentioned.

If a headline reads *“Bank of Spain: Data Points To Q2 GDP Contraction”*, the system automatically infers this story is about the country *“Spain”*. Since this story would match the event category *“gdp-guidance-down”* designed to match a country, the entity *“Spain”* would receive a relevance score of 100 and the entity *“Bank of Spain”* a score of 90 or above.

If an entity is identified in a *“source”* role, then it’s given a lower score of 10. A source may be a publisher, data provider, or firm that authored, originated, or is referenced in the story. However, if an entity is identified in a source role but also detected as a non-source role within the story, then the source role is disregarded (for the purpose of computing relevance), and it’s treated the same as any other entity described above.

The classifier that detects entities has time-sensitive information about each entity including short names, long names, abbreviations, security identifiers, subsidiary information, and up-to-date corporate actions data. This allows for *“point-in-time”* detection of entities in the text.

EVENT_SENTIMENT_SCORE

A granular score between -1.00 and +1.00 with 2 decimal places that represents the news sentiment for a given entity by measuring various proxies sampled from the news. The score is determined by systematically matching stories typically categorized by financial experts as having short-term positive or negative financial or economic impact. The strength of the score was determined by asking financial experts to classify entity-specific events in order to determine whether these events generally convey positive or negative sentiment and to what degree. Their ratings are encapsulated in an algorithm that generates a range for each event within -1.00 to +1.00 where 0 indicates neutral sentiment, positive values indicate positive sentiment, and negative values indicate negative sentiment.

EVENT_SENTIMENT_SCORE (ESS) probes many different sentiment proxies typically reported in financial news and categorized by RavenPack. The algorithm produces a score for more than 6,700 categories of business, economic, and geopolitical events ranging from earnings announcements to terrorist attacks. The score is determined by systematically detecting

entities and the roles played by those entities in a story using RavenPack’s proprietary technology and extensive database of time sensitive information about entities. The algorithms then dynamically assign an ESS score based on consensus-based score ranges and by performing analysis and computation when factors such as magnitudes, casualties, words that convey sentiment, comparative values, or ratings are disclosed in the story.

For example, the algorithm is capable of interpreting actual figures, estimates, ratings, revisions, magnitudes, and recommendations disclosed in news stories. It can compare actual versus estimated figures about earnings, revenues, or dividends and produce an ESS score based on the comparisons. It calculates percentage differences between financial figures and identifies and interprets stock and credit ratings disclosed by analysts.

The ESS algorithms can factor information such as the Richter scale in the case of an earthquake or the number of casualties in a suicide bombing event. The use of emotionally charged language by authors is also factored when shaping the strength component of the ESS.

The ESS algorithm has embedded information on rating scales from all major brokerage firms, investment banks, and credit rating agencies. It uses this information to differentiate and assess the various actions taken by analysts. For example, the algorithm generates a lower (more negative) ESS score for stories about an analyst downgrade from a “Strong Buy to a Strong Sell” than from a “Buy to a Neutral”. In the case of stories about financial results or economic indicators, it computes the percentage change between the disclosed actual figures versus the street consensus or any other benchmarks disclosed in the story. For example, a company beating earnings by 70% will receive a higher (more positive) ESS score than a company exceeding a benchmark by 1%.

ESS leverages RavenPack’s event detection technology and produces a sentiment score every time an event is matched. See the section on the Event Taxonomy for more information on Factors in the Event Sentiment Score.

EVENT_RELEVANCE

An integer between 0-100 that reflects the relevance of the event in the story. An event relevance score is assigned for records relating to an event, where related records can be identified by having the same EVENT_SIMILARITY_KEY within a story. The score is based on the earliest mention and frequency of the event match, according to the table below. The score is incremented for additional mentions of the event in the same paragraph, up to the maximum score for that paragraph.

Position	Event Relevance
First in Headline	100
Subsequent in Headline	90-99
Paragraphs 1 & 2	80-89

Rest of story body	0-79
--------------------	------

EVENT_SIMILARITY_KEY

A unique 32 character key that identifies similar events in the RavenPack Analytics data. All stories with similar events across the entire archive and those arriving on the real-time feed share the same similarity key. Events are considered to be similar when they are categorized with all of the same attributes pertaining to an event, such as the roles each entity plays and percentage changes, magnitudes etc. captured.

EVENT_SIMILARITY_DAYS

A granular number with up to 5 decimal places which indicates the number of days since a story with a similar event was detected over the last 365 days. Values range between 0.00000 and 365 inclusive. A value of 365 means that the most recent story with a similar event may have occurred 365 or more days in the past. The value 0.00000 means a story with a similar event occurred at the exact same timestamp.

TOPIC

A subject or theme of events detected by RavenPack. The highest level of the RavenPack Event Taxonomy.

GROUP

A collection of related events. The second highest level of the RavenPack Event Taxonomy.

TYPE

A class of events, the constituents of which share similar characteristics.

SUB_TYPE

A subdivision of a particular class of events.

PROPERTY

A generic named attribute of an event such as an entity, role, or string extracted from a matched event type. When applicable, the role played by the entity in the event is detected and tagged.

FACT_LEVEL

An indicator of how close the event relates to a fact. Values are:

Fact Level	Description
fact	An event that came from a concrete statement of information.
forecast	An event that gives guidance about the future.

opinion	An event that expresses a view or a hypothesis about a statement of information.
---------	--

RP_POSITION_ID

A unique and permanent identifier for positions assigned by RavenPack. Every position tracked is assigned a unique entity identifier comprised of 6 alphanumeric characters. A full list of RP_POSITION_IDs is available via the Entity Mapping API.

POSITION_NAME

The position held by an individual within the entity involved in a specific news event. A full list of POSITION_NAMEs is available via the Entity Mapping API.

EVALUATION_METHOD

A period of time used to measure changes from previous levels in an event. Currently RavenPack supports the following evaluation methods as named attributes:

Method	Description
YOY	Year-on-Year change
QOQ	Quarter-on-Quarter change
MOM	Month-on-Month change
LFL	Like-for-Like comparison

MATURITY

For events related to debt, this named attribute indicates the period of time for which a financial instrument remains outstanding. The time period is represented by the following formats:

Format	Description
{1-365}-DAY	Maturity in days. Prefix is a number from 1 to 365, e.g. 2-DAY.
{1-52}-WK	Maturity in weeks. Prefix is a number from 1 to 52, e.g. 4-WK.
{1-12}-MTH	Maturity in months. Prefix is a number from 1 to 12, e.g. 9-MTH.
{1-50}-YR	Maturity in years. Prefix is a number from 1 to 50, e.g. 40-YR.

EARNINGS_TYPE

For events about earnings (Net Income, EPS, EBITDA, etc.), this named attribute indicates the type of calculation method used for reporting. RavenPack supports the following types of calculations. An empty value means that the calculation method is unspecified.

Type	Description
reported	Represents GAAP EPS for all U.S. companies. Includes exceptionals, nonrecurring items, and stock option expense (SOE). Since analysts often provide both adjusted and non-adjusted EPS figures, this distinguishes itself as the non-adjusted figure.
non-gaap	Not according to GAAP. These values do not include non-recurring items and include SOE if it has been reported by the company.
ex-exceptionals	Fully adjusted (non-GAAP) to exclude both extraordinary items and SOE.
adjusted	Represents non-GAAP EPS for all U.S. companies. Excludes exceptionals, nonrecurring items, and SOE.
non-diluted	Accounts for all the P&L from operational, trading, and interest activities, that have been discontinued or acquired at any point during the year. Excludes any profit or loss associated with the sale or termination of discontinued operations, fixed assets or related businesses, or from any permanent devaluation or write off of their values. Does not factor in the dilutive effects on convertible securities.
diluted-adjusted	Labelled in brokers' research report as being adjusted for any non-recurring, discontinued operations, and/or exceptional items. Based on the diluted shares.
diluted-reported	Represents GAAP earnings "as reported", calculated by the analysts in accordance with the accounting standards by which the company abides. Based on diluted shares and includes all extraordinary and unusual items.
headline-basic	Accounts for all the P&L from operational, trading, and interest activities that have been discontinued or acquired at any point during the year. Excludes any profit or loss associated with the sale or termination of discontinued operations, fixed assets or related businesses, or from any permanent devaluation or write-off of their values. Does not factor in the dilutive effects on convertible securities.
headline-diluted	Accounts for all the P&L from operational, trading, and interest activities that have been discontinued or acquired at any point during the year. Excludes any profit or loss associated with the sale or termination of discontinued operations, fixed assets or related businesses, or from any permanent devaluation or write-off of their values. Calculated using fully diluted shares outstanding.
consolidated	Data for a given company is merged with all of its affiliates and is consolidated.
standalone	Represents the data for the parent company. Also known as Non-Consolidated.

EVENT_START_DATE_UTC

The date when the event starts in UTC. For example, in a story received at 2016-09-01 13:00:00 that says: "Microsoft will host a conference call on September 2nd", the `TIMESTAMP_UTC` will be 2016-09-01 13:00:00.000 and the `EVENT_START_DATE_UTC` will be 2016-09-02 00:00:00.

EVENT_END_DATE_UTC

The date when the event ends in UTC. For example, in a story received at 2016-09-01 13:00:00 that says: "Microsoft will host a conference call on September 2nd", the `TIMESTAMP_UTC` will be 2016-09-01 13:00:00.000 and the `EVENT_END_DATE_UTC` will be 2016-09-03 00:00:00.

REPORTING_PERIOD

A period on a financial calendar that acts as a basis for reporting business information.

Format	Description
YYYY-Q{1-4}	Quarter-long regular calendar period, e.g. 2016-Q1
FY-YYYY-Q{1-4}	Quarter-long fiscal calendar period, e.g. FY-2016-Q1
YYYY-H{1-2}	Half-year regular calendar period, e.g. 2016-H1
FY-YYYY-H{1-2}	Half-year fiscal calendar period, e.g. FY-2016-H1
YYYY-9MTH	9-Month regular calendar period, e.g. 2016-9MTH
FY-YYYY-9MTH	9-Month fiscal calendar period, e.g. FY-2016-9MTH
YYYY	Year-long regular calendar period e.g. 2016
FY-YYYY	Year-long corporate fiscal calendar period e.g. FY-2016

REPORTING_START_DATE_UTC

The start of the reporting period associated with the event. For example, in a story received at 2016-08-01 13:00:00 that says: "Employment Advances in Q2", the `TIMESTAMP_UTC` will be 2016-08-01 13:00:00.000 and the `REPORTING_START_DATE_UTC` will be 2016-04-01 00:00:00.

REPORTING_END_DATE_UTC

The end of the reporting period associated with the event. For example, in a story received at 2016-08-01 13:00:00 that says: "Employment Advances in Q2", the `TIMESTAMP_UTC` will be 2016-08-01 13:00:00.000 and the `REPORTING_END_DATE_UTC` will be 2016-07-01 00:00:00.

RELATED_ENTITY

An entity that relates directly to another entity within the context of the same event in a story. Within the story, the `EVENT_SIMILARITY_KEY` will be the same for both records. See the `RELATIONSHIP` field for an indicator of the type of relationship.

RELATIONSHIP

The type of relationship between entities reported in the `RELATED_ENTITY` field. Values are:

Relationship	Description
--------------	-------------

PRODUCT	A product that is owned by a company.
OWNER	The company that owns a product.

CATEGORY

A unique tag to label, identify, and recognize a particular type and property of an entity-specific news event. The category is a canonical version of the event that has been captured.

EVENT_TEXT

A short text summary of an event captured in a document.

NEWS_TYPE

Classifies the type of news story into one of the following categories:

Type	Description
HOT-NEWS-FLASH	Headline and no body text, marked as breaking news in the editorial process.
NEWS-FLASH	Headline and no body text.
FULL-ARTICLE	Headline and one or more paragraphs of mostly textual material.
PRESS-RELEASE	Corporate announcement created by an entity and distributed by a news wire.
TABULAR-MATERIAL	Headline and one or more segments of mostly tabular data.
RNS-SEC8K	A news article that came from an SEC 8K filing.
RNS-SEC10K	A news article that came from an SEC 10K filing.
RNS-SEC10Q	A news article that came from an SEC 10Q filing.
RNS-SEC13D	A news article that came from an SEC 13D filing.
RNS-SEC13F	A news article that came from an SEC 13F filing.
RNS-SEC144	A news article that came from an SEC 144 filing.

RP_SOURCE_ID

A unique and permanent news source identifier assigned by RavenPack. Every news provider tracked is assigned a unique identifier comprised of 6 alphanumeric characters.

SOURCE_NAME

The official canonical name for the news source.

CSS – COMPOSITE SENTIMENT SCORE

A granular score between -1.00 and +1.00 with up to 2 decimal places that represents the news sentiment of a given story by combining various sentiment analysis techniques. The direction of the score is determined by looking at emotionally charged words and phrases and by matching stories typically rated by experts as having short-term positive or negative share price impact. The strength of the score (values above or below 0, where 0 represents neutral strength) is determined from intraday stock price reactions modeled empirically using tick data from approximately 100 large cap stocks.

CSS combines 5 sentiment analytics (PEQ, BEE, BMQ, BAM, and BCA) using an intuitive set of rules while ensuring no sentiment disagreement exists amongst the analytics. One way of using CSS scores may involve rules like:

```
If CSS > 0 Then ' (Positive Signal)
    ' Go long
Elseif CSS < 0 Then ' (Negative Signal)
    ' Go short
Elseif CSS = 0 Then ' (Neutral Signal)
    ' Hold position
End If
```

CSS was trained on market data using a portfolio of large cap stocks and evaluating intraday fluctuations to determine “strength” or how positive or negative a story is. Using the “strength” aspect of this score may depend on your investment strategy and trading horizon. Typically, CSS scores hover between [-0.2, +0.2] so higher or lower values are assigned only in cases where confidence is high on short term signals. This score combines RavenPack's Traditional, Expert Consensus, and Market Response methodologies.

NIP – NEWS IMPACT PROJECTIONS

A granular score between -1.00 and 1.00 with up to 2 decimal places that represents the degree of impact a news flash has on the market over the following two-hour period. The training set for this classifier used tick data for a test set of large cap companies and looked at the relative volatility of each stock price measured in the two hours following a news flash. The relative volatility is the volatility divided by the mean of volatilities of all companies in the test set during the same period. The classifier is trained to predict whether relative volatility is high or low given the language used by journalists in news flashes, typically about corporate actions and analyst revisions.

Whether something is considered high or low depends on the time of day when the story arrived. The score is centered at 0, which represents zero impact; values above 0 indicate higher impact in terms of volatility. The more extreme the impact value, the higher the confidence of the score. Scores below 0 indicate low or unknown impact and lower confidence in the score.

The best performance of the score is obtained when filtering for RELEVANCE above 90. This NIP score is based on RavenPack's Market Response Methodology.

PEQ – GLOBAL EQUITIES

A score that represents the news sentiment of the given news item according to the PEQ classifier, which specializes in identifying positive and negative words and phrases in articles about global equities. Scores can take values of -1, 0, or 1 indicating negative, neutral, or positive sentiment, respectively. This sentiment score is based on RavenPack's Traditional Methodology.

BEE – EARNINGS EVALUATIONS

A score that represents the news sentiment of the given story according to the BEE classifier, which specializes in news stories about earnings evaluations. Scores can take values of -1, 0, or 1 indicating negative, neutral, or positive sentiment, respectively. This sentiment score is based on RavenPack's Traditional Methodology.

BMQ – EDITORIALS & COMMENTARY

A score that represents the news sentiment of the given story according to the BMQ classifier, which specializes in short commentary and editorials on global equity markets. Scores can take values of -1, 0, or 1 indicating negative, neutral, or positive sentiment, respectively. This sentiment score is based on RavenPack's Expert Consensus Methodology.

BAM – VENTURE, COMPANY, MERGERS & ACQUISITIONS

A score that represents the news sentiment of the given story according to the BAM classifier, which specializes in news stories about mergers, acquisitions, and takeovers. Scores can take values of -1, 0, or 1 indicating negative, neutral, or positive sentiment, respectively. This sentiment score is based on RavenPack's Expert Consensus Methodology, and has been trained on stories that lead up to pre-identified mergers, acquisitions, and takeover events.

BCA – REPORTS ON CORPORATE ACTIONS

A score that represents the news sentiment of the given news story according to the BCA classifier, which specializes in reports on corporate action announcements. Scores can take values of -1, 0, or 1 indicating negative, neutral, or positive sentiment, respectively. This sentiment score is based on RavenPack's Expert Consensus Methodology and has been trained on stories that lead up to a pre-identified corporate action announcement.

BER – EARNINGS RELEASES

A score that represents the news sentiment of the given story according to the BER classifier, which specializes in news stories about earnings releases. Scores can take values of -1, 0, or 1 indicating negative, neutral, or positive sentiment, respectively. This sentiment score is based on RavenPack's Expert Consensus Methodology.

ANL_CHG – ANALYST RECOMMENDATIONS & CHANGES

A score that represents a change in recommendation by an analyst firm in the form of a numerical score. When the mention of a company in a story matches the criteria for ANL-CHG, scores can take values of -1, 0, or 1, indicating a downgrade, neutral, or upgrade rating, depending on the recommendation change by the analyst. This analytic is based on the event category results of a) analyst-ratings-change-positive and b) analyst-ratings-change-negative. This analytic contains values only for news stories that disclose changes in analyst recommendations.

MCQ – MULTI CLASSIFIER FOR EQUITIES

A score that represents the news sentiment based on the tone; applicable only towards the most relevant entities mentioned in a story. The score is derived from a combination of analytics values produced by the BMQ, BEE, BCA, and ANL-CHG classifiers. An MCQ score can have a non-neutral value when the relevance score for an entity is 90 or higher and either there is an ANL-CHG score or all of BMQ, BEE, and BCA scores are positive (1) and neutral (0) or negative (-1) and neutral (0). Entities with a relevance of less than 90 will always have a neutral score. The logic behind this analytic is to detect consistent sentiment classifications, discarding combinations where these classifiers may have contradictory scores. MCQ scores can take values of -1, 0, or 1 indicating negative, neutral, or positive sentiment, respectively.

RP_STORY_EVENT_INDEX

An integer representing the sequence in which each record is presented within a document. This integer must be less than or equal to the RP_STORY_EVENT_COUNT.

RP_STORY_EVENT_COUNT

Represents the total entity records published by RavenPack per news story.

PROVIDER_ID

Identifies the provider of the content. RavenPack currently receives news from the following providers and aggregators:

Provider	Description
AN	Alliance News
BZG	Benzinga Pro
DJ	Dow Jones Newswires and Third Party Content Wires
FLY	The Fly
FXS	FX Street News and FX Street Economic Calendar
MRVR	MoreOver News and Social Media

MT	Midnight Trader News and Midnight Trader MtPro
RP	RavenPack-originated content

PROVIDER_STORY_ID

An alphanumeric character identifier to uniquely identify the news story in the provider's universe. Here is a key to help interpret these identifiers:

Provider	Format	Example
AN	<id-string>	1420095499035228600
BZG	<bz:id> + ":" + <bz:revisionId>	123123:1232323
DJ	Product + Docdate + Seq	DN20010228008585
FLY	<fly-unique-id>	2122561
FXS News	NEWS + ":" + FXS_storyId	NEWS:bde3b58a-88d2-4a03-8c4e-98d4e9da28a
FXS EcoCal	ECO + ":" + idEcoCalendarDate	ECO:21d28292-d250-4b94-b5ad-e88cfb02ea62
MRVR	FeedVersion + ":" + ArticleId	10:16481270273
MT	TransmissionID	A754249
MT Pro	Filename	form_01152014_2377436.xml
RP	<rp_story_id>	1FB2B3F5E99C4D3BCF59FDB3E8C8C9BD

HEADLINE

The headline or summary text for the document.

PRODUCT_KEY

Identifies which content set the record came from. For RavenPack Analytics, the product key is always "RPA".

Functions (Indicators)

The RavenPack Platform allows you to create your own custom daily functions (indicators) based on RavenPack data, which can be used as signals to power trading models.

Daily data for custom functions is aggregated on a per entity basis into daily buckets such that a datafile will have a set of values per day for each entity in the dataset, with each value

representing the information up to the time of the record. In addition to values per day for each entity, a ROLLUP record is produced that aggregates data across all entities in the dataset per day. When requesting data, it is possible to request daily data sampled at any time of the day by specifying the desired time in the start_date. For example, if you select a daily dataset with a start_date of 2017-01-02 19:00 UTC, each datapoint will represent information up to 19:00 each day.

The number of functions that may be included in a dataset or in an API request is limited to 30 for all dataset, datafile, and JSON requests.

Supported Functions

One may describe daily aggregates of the data by specifying functions within custom fields, which should contain the functions used for aggregating. These functions run on either granular data-fields, or on other previously defined custom fields by the user in the same dataset.

The list of functions that may be used for specifying custom fields are as follows:

avg

Average of the values in a given field over the entire timeframe specified. Nulls are excluded. Applies only to numerical fields.

buzz

A function that compares the daily volume with the average volume during the period specified. May be calculated only on RP_ENTITY_ID and CATEGORY. The formula that describes this function is:

$$Buzz = \frac{News_Volume_1D - avg(News_Volume_1D)}{sd(News_Volume_1D)}$$

The News_Volume_1D is the daily count of the values in the field specified. The avg(News_Volume_1D) and sd(News_Volume_1D) capture the mean and standard deviation of News_Volume_1D over the window specified.

count

Count of the values for a field for each day. Nulls are excluded. Applies only to RP_ENTITY_ID and CATEGORY.

cardinality

Count of the distinct values for a field for each day. Nulls are excluded. Cardinality does not support a window size. Applies to the fields RP_ENTITY_ID, RP_STORY_ID and CATEGORY.

daily_avg

Average of all of the daily averages over the timeframe specified. Nulls are excluded. Applies only to numerical fields.

min and max

The minimum and maximum value of a field over the timeframe specified. Applies only to numerical fields.

stddev

Standard deviation of the values over the day. Nulls are excluded. Applies only to numerical fields.

strength

An indicator which incorporates a decay function to give more weight to recent news. Applies only to numerical fields. In order to calculate the strength of an analytic field, first we have to calculate the daily average value. For example if we wanted to calculate the daily average value for the Event Sentiment Score we would use the formula below:

$$AVERAGE_1D = \sum_{i=1}^n \frac{EventSentimentScore_i}{n}, \text{ for } i \in U$$

Where $U = \{1, \dots, n\}$ is the number of news events for the company over the past 24 hours. Then we would apply the next formula below to calculate Strength.

$$Strength(LB) = \sum_{i=0}^{LB-1} \frac{w_{1,i} * w_{2,i} * AVERAGE_1D_i}{\sum_{j=0}^{LB-1} IF(AVERAGE_1D_j = NULL) THEN 0 ELSE w_{2,j}}$$

Where LB is the lookback window and w_1 and w_2 are both exponential functions with different speed of decay. However, they take two different roles:

- $w_{1,i} = 2^{(-d_i/LB)}$ with $d_i = 0 \dots LB-1$ where d_i represents the number of days since now. w_1 is used to degrade the importance of each event based on the elapsed time, i.e. representing a degradation function.
- $w_{2,i} = 2^{(-10*d_i/LB)}$ with $d_i = 0 \dots LB-1$ producing a set of weights placing more importance on the most recent news – thereby increasing the speed of decay for older news. In addition, w_2 ensures that the sentiment score is bounded between -1 and 1.

Effectively, by including both w_1 and w_2 , we allow for an asymmetric weighting function that decays news faster for high news volume stocks (large cap) than for low news volume stocks (small cap). Take for example a company that has only one news event during a 91-day window. In such case, only w_1 will be "in use", as w_2 will only contribute with one coefficient that will cancel out by the normalization term. For a company with more than one news event, some weight will be placed on all of them. However, it will not be the same. The further apart in time, the more weight will be placed on the most recent event – increasing the speed of decay for all other news.

sum

Sum of all the values for a field for each day. Applies only to numerical fields.

rank

Ranks each entity in a dataset for each day in order of the field specified. Fields with equal values for the ranking criteria receive the same rank. We then add the number of tied rows to the tied rank to calculate the next rank. Therefore, the ranks may not be consecutive numbers. This function requires a user defined field as the input field.

dense_rank

Consecutively ranks each entity in a dataset for each day in order of the field specified. The ranks are consecutive integers beginning with 1. The largest rank value is the number of unique values returned by the query. Rank values are not skipped in the event of ties. Rows with equal values for the ranking criteria receive the same rank. This function requires a user defined field as the input field.

row_number

Returns the sequential number of a row for each day in order of the field specified. Ties are broken by RP_ENTITY_ID. This function requires a user defined field as the input field.

ntile:

Sorts the daily set of entities by "order by" field and divides the list of entities equally into a specified number of buckets. The Ntile is the index of each bucket in the resulting set.

Buckets are the number of buckets that rows will be divided into (Valid input: 1-100). This function requires a user defined field as the input field.

Indicator fields may be specified either via the RavenPack Platform or by using the RavenPack Web API. For more details on syntax for functions in custom fields, please refer to the API Documentation.

Document Search Filters

When using the RavenPack Platform, you may further refine the data by searching for keywords within the underlying documents. The syntax supported by the document search filter allows for a number of different operations, including the use of logical operators, grouping of expressions, automatic-stemming and fixed-text searches.

For further reference on search operators and order of precedence, please refer to the [Document Search Syntax](#) guide within the API Documentation.

Alerts

Alerts are a great way to stay on top of market changes while you are away from your desk. The Ravenpack Platform allows users to create alerts and to be notified when specific criteria occur.

It is possible to create alerts on any dataset and specify flexible conditions that focus on a specific set of entities, list of events or analytics such as relevance, novelty and sentiment. It is also possible to create alerts based on indicators such that notifications are sent when aggregate measures such as a moving average on sentiment cross over a specified threshold.

Alerts are currently designed to integrate directly with Email and Slack and can be used to notify either individual users or be broadcast to an entire distribution list or channel.

To find out more about Alerts, please contact your RavenPack sales representative.

RavenPack Reference Service

RavenPack supplies a proprietary reference data service for all entities it tracks and monitors in the media. The service indicates the links and relationships between entities in business, economic, and political structures within a geographic space. Entities include companies (stocks), organizations, geographical locations, currencies, and commodities. The entity-reference service is provided via the entity-reference API described in our API documentation. It is also possible to automatically map entities using the entity mapping API or interactively when creating a dataset on the RavenPack platform.

Entities tracked by RavenPack are keyed using the RP_ENTITY_ID and for each entity the mapping file lists identifying information such as names and securities along with the validity dates for each value. The reference service allows proprietary universes of entities (e.g. securities) to be cross-referenced with the hundreds of thousands of entities tracked by RavenPack over time.

The RavenPack Reference Service also provides unique relationships between entities. For example, it maps countries to membership organizations such as the European Union, the Euro-zone, G20, OPEC, OECD, and NATO or to trading blocs such as NAFTA, MERCOSUR, or The League of Arab States, among others. Relationships are point-in-time sensitive in that it keeps track of when entities join (or leave) any given membership. By delineating a country to a supra-organization or trading bloc for example, at any point in time, this mapping data helps expedite and simplify macroeconomic and geopolitical analysis by clients.

In addition, RavenPack maps all currencies to their corresponding countries and governments. In the case of multi-government denominated currencies such as the Euro, we map it only to the member states of the European Union that form part of the Eurozone.

Reference Data File Contents

Data Format:	Comma separated values (.csv) files
File Encoding:	Latin-1
Data Fields:	6 fields

The key fields available in the entity reference service are:

RP_ENTITY_ID

A permanent 6 character key assigned by RavenPack for uniquely identifying entities over time. This should be used for linking any data in the analytics data files.

ENTITY_TYPE

The type of entity associated with a particular RP_ENTITY_ID. Currently RavenPack supports the following entity types:

Type	Name	Description
CMDT	Commodity	Exchange traded commodities such as crude oil and soy.
COMP	Company	Business organizations that may be traded on a stock exchange.
CURR	Currency	Currencies of all financial and industrial countries.
NATL	Nationality	The status of belonging to a particular nation.
ORGA	Organization	Non-business organizations, e.g. governments and central banks.
ORGT	Organization Type	A categorization of related organizations.
PEOP	People	Individuals that are mentioned in the news.
PLCE	Place	Towns, cities, and countries.
PROD	Product	Products and services.
PRDT	Product Type	A categorization of related products.
TEAM	Sports Teams	Professional teams from a variety of different sports.
POSI	Position	A position held by an individual within an entity.
SRCE	Source	A source of content from which analytics are derived.

NOTE: To provide as many details as possible about entities, new data types may appear as and when they are added to our database. We advise that any code used to parse this document should allow for these unspecified values.

DATA_TYPE

Categorizes 26 types of data relevant to this version of the product:

Type	Description
COMPANY	A company identifier comprised of a code based on the company's country of incorporation and a local exchange ticker or symbol. This is used primarily for visual representation in the data files and will change over time.

COUNTRY	A country associated with this entity.
COUNTRY_ID	A RP_ENTITY_ID of the entity of the country associated with this entity.
CUSIP	A known CUSIP identifier for a stock of the detected company.
DESCRIPTION	A description of the entity where applicable.
ENTITY_NAME	The official canonical name for the entity as listed in the data files.
GEONAMEID	For a place, this is the geoname from the geonames data source.
GOVERNMENT	For a place, this is the associated central government organization.
HAS_MEMBER_ID	Indicates that the RP_ENTITY_ID listed in the DATA_VALUE field is a constituent or member of this entity.
ISIN	A known ISIN (International Securities Identifier Number) for the company.
IS_MEMBER_ID	Indicates that this entity is a constituent or member of the RP_ENTITY_ID listed in the DATA_VALUE field.
ISO_CODE	Defines the two-letter ISO country code for the country.
LATITUDE	A geographic coordinate for latitude of the place represented in decimal degrees (DD).
LISTING	A combination of the MIC and a local exchange ticker/symbol for the company. For example, the NASDAQ ticker for Apple Inc. would be represented as XNAS:APPL.
LONGITUDE	A geographic coordinate for longitude of the place represented in decimal degrees (DD).
MIC	The ISO 10383 Market Identifier Code (MIC) is a four-character code used to identify stock markets and other trading exchanges within global trading and referencing systems.
NAME	The official registered name for the entity.
NATIONALITY	The nationality of a person entity.
ORGANIZATION_TYPE	The type of ORGANIZATION associated with the entity.
ORGANIZATION_TYPE_ID	The unique ENTITY_ID of the ORGANIZATION_TYPE.
PARENT_ORG_ID	Indicates that this ORGANIZATION is a child of the ORGANIZATION whose RP_ENTITY_ID is listed in the DATA_VALUE field.
PARENT_PRODUCT_TYPE	Product types form a hierarchy, e.g. Electric Vehicle is a type of Automobile. The PARENT_PRODUCT_TYPE is the RP_ENTITY_ID of the parent product type in the hierarchy.

PARENT_SOURCE	Indicates that this Source entity is a child of the Source entity listed in the DATA_VALUE field.																
PLACE_TYPE	The type of PLACE. This can be one of COUNTRY, CITY, REGION_1 or REGION_2. The REGION_2 and REGION_1 fields can vary according to the regional differences within the country. For example, in the US, REGION_2 will be a county and REGION_1 a state. However, in Spain, REGION_2 will be a province and REGION_1 an autonomous region, etc.																
PRODUCT_OWNER	The RP_ENTITY_ID of the company that owns a particular product. Products may have multiple owners.																
PRODUCT_TYPE	The type associated with a particular product.																
PROVINCE	For a place, this is the region where the place is found.																
PUBLICATION_TYPE	Identifies the general type of the news for a Source entity. Its value can be one of the following: BLOG: A discussion or informational website consisting of discrete entries or postings. NEWS: A source providing up-to-the-minute news stories, financial market updates, and other newly received or noteworthy information. JOURNAL: A scholarly publication containing articles written by researchers, professors and other experts. OTHER: A variety of other information sources beyond blogs or news providers. Sources can include company websites, government and regulatory agencies, among others.																
REGION_ID	For Places, the REGION_ID is the RP_ENTITY_ID of the region in which the Place entity is located.																
SEDOL	A known SEDOL identifier for a stock of the detected company.																
SOURCE_RANK	For a Source entity, this is a categorization of influence and trustworthiness of a news provider. The ranking is based on a range from 1 to 10 where rank 1 is the highest (i.e. reliable and fully accountable) and 10 is the lowest (i.e. unverified or potentially fabricated). Below are the descriptions of the 10 possible ranks: <table border="1" data-bbox="578 1520 1399 1887"> <thead> <tr> <th>Rank</th> <th>News Source Classification</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>Fully accountable, reputable and balanced</td> </tr> <tr> <td>2</td> <td>Official, reliable and honest</td> </tr> <tr> <td>3</td> <td>Acknowledged, formal, and credible</td> </tr> <tr> <td>4</td> <td>Known and reasonable credibility</td> </tr> <tr> <td>5</td> <td>Satisfactory credibility</td> </tr> <tr> <td>6</td> <td>Limited influence and sincerity</td> </tr> <tr> <td>7</td> <td>Biased and seemingly unreliable</td> </tr> </tbody> </table>	Rank	News Source Classification	1	Fully accountable, reputable and balanced	2	Official, reliable and honest	3	Acknowledged, formal, and credible	4	Known and reasonable credibility	5	Satisfactory credibility	6	Limited influence and sincerity	7	Biased and seemingly unreliable
Rank	News Source Classification																
1	Fully accountable, reputable and balanced																
2	Official, reliable and honest																
3	Acknowledged, formal, and credible																
4	Known and reasonable credibility																
5	Satisfactory credibility																
6	Limited influence and sincerity																
7	Biased and seemingly unreliable																

	8 Unknown identity, strongly biased, and unconvincing 9 Illusory and misleading 10 Unverifiable, fake, and fabricating
SYMBOL	For a currency, this is the associated symbol.
TEAM_TYPE	The type of sport that the team is involved in.
TICKER	The local exchange ticker or symbol for the company.
TYPE	A tag that classifies a particular ORGANIZATION_TYPE, e.g. GOVERNMENT.
VALIDITY	Indicates the period of time where a news source is actively providing coverage.

DATA_VALUE

The value of a particular piece of information during the time indicated by the RANGE_START and RANGE_END.

RANGE_START

A DATA_VALUE is valid from this date. An empty RANGE_START means that the DATA_VALUE has no start date and was valid from or before the year 2000.

RANGE_END

A DATA_VALUE is valid until this date. An empty RANGE_END means that the DATA_VALUE has no end date and is thus still valid.

The RavenPack Taxonomy

The RavenPack Taxonomy is a comprehensive structure for content classification. It provides a definitive system categorizing structured and unstructured information, enabling analysis on thousands of entities including companies, products, people, organizations, places, and more. The RavenPack Taxonomy is supported by the RavenPack Entity Database, an unrivalled data source of named entities that is constantly kept up to date.

Classification Structure

The taxonomy enables content analysis across four levels of classification. It offers a balance between levels of information aggregation, for those who look at information from the top down, and granularity, for those who look at specific events from the bottom up.

TOPIC A subject or theme of events. The highest level of the RavenPack Taxonomy.

GROUP A collection of related events.

TYPE A class of events, the constituents of which share similar characteristics.

SUB_TYPE A subdivision of a particular class of events.

PROPERTY A named attribute of an event such as a role.

CATEGORY A tag to label a particular type and property of an event.

Properties

Each type of event may contain a series of properties that provide more context of what is being classified. Published properties include things like:

Property	Description
Roles	Raters, plaintiffs, defendants, suppliers, target products, locations, etc.
Positions	CEO, CTO, Director, Musician, etc.
Temporal data	Dates, reporting periods, and maturity dates.
Indicators	Bearish, bullish, overbought, oversold, etc.
Benchmarks	Evaluation methods like Yr-on-Yr or earnings types like Diluted-Adjusted.
Opinions	Views on companies, products, people, etc.

Many properties, like roles, are captured in the generic PROPERTY field, while more specialized ones, like the evaluation method of an indicator, the maturity of a debt issuance, or the type of earnings calculation used, are recorded in their own unique fields.

Fact Level

The Fact Level identifies whether an event is written as something that is known or proved to be true, or as a view or judgement, not necessarily based on fact or knowledge. Every matching event in the RavenPack taxonomy contains a *Fact*, a *Forecast* or an *Opinion* tag. If a matching event is tagged as a *Fact*, this means the underlying information was published as if something has really occurred or is actually the case. When an event is related to future guidance, the *Forecast* tag is used. When the *Opinion* tag is used, the information reads as a belief, judgement, or viewpoint on a topic level in the RavenPack taxonomy.

Categories

Every event that contains a primary role or distinguishable attribute is categorized with a unique tag called a Category. These tags normally contain the name of the event followed by the role played by the matching entity. If there is only one primary role in the event, key attributes such as indicators (bearish, bullish, etc.) may be used as part of the category name to make it unique.

The Benefits

- **Business, economic, & geopolitical focus** - RavenPack offers increased clarity, better structure, and universality to meet the needs of various industries including business, finance, media, and government.
- **Comprehensive coverage** - RavenPack covers events affecting most companies, products, organizations, places, currencies, and commodities that a business or financial professional will care about.
- **Accurate and timely** - The accuracy, quality, and timeliness of the data make it a very reliable reference for researchers and analysts.
- **Efficient and effective** - A rigorous and transparent rule-based methodology enables accurate and relevant classification. RavenPack is a single data source, removing the need for users to maintain individual databases or undertake costly and time-consuming tasks to develop and maintain their own taxonomies.

The Event Taxonomy File

The RavenPack taxonomy can be downloaded [here](#).

The RavenPack taxonomy file includes the following elements:

TOPIC

A subject or theme of events. The highest level of the RavenPack Taxonomy.

GROUP

A collection of related events.

TYPE

A class of events, the constituents of which share similar characteristics.

SUB_TYPE

A subdivision of a particular class of events.

PROPERTY

A named attribute of an event such as an entity, role, number, or string extracted from a matched event type.

FACT_LEVEL

A tag identifying the event as either a fact, a forecast, or an opinion.

CATEGORY

A tag to label, identify, or recognize a particular type and property of an event.

DESCRIPTION

An explanation describing the meaning of the category tag and the context where it can match.

SCHEDULED

If the announcement of the event was anticipated, arranged, or planned according to some schedule or timetable, then the value will be TRUE otherwise it will be FALSE.

VALID_ENTITY_TYPES

The ENTITY_TYPE elements, which are valid matches for that particular category.

The RavenPack Taxonomy can be downloaded in .csv file format from the Product Area.

Dataset Schemas

This section explains the data types and maximum lengths for the fields included in the RPA Datasets.

TIMESTAMP_UTC

The Date/Time (YYYY-MM-DD HH24:MI:SS.FF3) at which the news item was received by RavenPack servers in Coordinated Universal Time (UTC). This field cannot contain a null value.

TIMESTAMP_TZ

The Date/Time (YYYY-MM-DD HH24:MI:SS.FF3) at which the news item was received by RavenPack servers in the time zone specified by the user. This field is returned whenever it is specified in the list of FIELDS and will return values in UTC by default when no TIME_ZONE is specified. This field cannot contain a null value.

RP_STORY_ID

A fixed-width character field of 32 characters.
Oracle example: CHAR(32). This field cannot contain a null value.

RP_ENTITY_ID

A fixed-width character field of 6 characters.
Oracle example: CHAR(6). This field cannot contain a null value.

ENTITY_TYPE

A fixed-width character field of 4 characters.
Oracle example: CHAR(4). This field cannot contain a null value.

ENTITY_NAME

A variable-width character field with a maximum of 400 characters.
Oracle example: VARCHAR2(400).

COUNTRY_CODE

A fixed-width character field of 2 characters.
Oracle example: CHAR(2).

RELEVANCE

A numerical field that stores an integer value from 0 to 100.

Oracle example: NUMBER(3,0). This field cannot contain a null value.

EVENT_SENTIMENT_SCORE

A numerical field that stores a floating point number value from -1.00 to +1.00 with up to 2 decimal places.

Oracle example: NUMBER(3,2).

EVENT_RELEVANCE

A numerical field that stores an integer value from 0 to 100.

Oracle example: NUMBER(3,0).

EVENT_SIMILARITY_KEY

A variable-width character field with a maximum of 32 characters.

Oracle example: VARCHAR2(32).

EVENT_SIMILARITY_DAYS

A numerical field that stores a floating point number value from 0 to 365 with up to 5 decimal places.

Oracle example: NUMBER(8,5).

TOPIC

A variable-width character field with a maximum of 50 characters.

Oracle example: VARCHAR2(50).

GROUP

A variable-width character field with a maximum of 50 characters.

Oracle example: VARCHAR2(50).

TYPE

A variable-width character field with a maximum of 50 characters.

Oracle example: VARCHAR2(50).

SUB_TYPE

A variable-width character field with a maximum of 50 characters.

Oracle example: VARCHAR2(50).

PROPERTY

A variable-width character field with a maximum of 50 characters.
Oracle example: VARCHAR2(50).

FACT_LEVEL

A variable-width character field with a maximum of 10 characters.
Oracle example: VARCHAR2(10).

RP_POSITION_ID

A fixed-width character field of 6 characters.
Oracle example: CHAR(6)

POSITION_NAME

A variable-width character field with a maximum of 400 characters.
Oracle example: VARCHAR2(400).

EVALUATION_METHOD

A variable-width character field with a maximum of 50 characters.
Oracle example: VARCHAR2(50).

MATURITY

A variable-width character field with a maximum of 50 characters.
Oracle example: VARCHAR2(50).

EARNINGS_TYPE

A variable-width character field with a maximum of 50 characters.
Oracle example: VARCHAR2(50).

EVENT_START_DATE_UTC

Date/Time (YYYY-MM-DD HH24:MI:SS) in Coordinated Universal Time (UTC).

EVENT_END_DATE_UTC

Date/Time (YYYY-MM-DD HH24:MI:SS) in Coordinated Universal Time (UTC).

REPORTING_PERIOD

A variable-width character field with a maximum of 50 characters.
Oracle example: VARCHAR2(50).

REPORTING_START_DATE_UTC

Date/Time (YYYY-MM-DD HH24:MI:SS) in Coordinated Universal Time (UTC).

REPORTING_END_DATE_UTC

Date/Time (YYYY-MM-DD HH24:MI:SS) in Coordinated Universal Time (UTC).

RELATED_ENTITY

A variable-width character field with a maximum of 6 characters.

Oracle example: VARCHAR2(6).

RELATIONSHIP

A variable-width character field with a maximum of 50 characters.

Oracle example: VARCHAR2(50).

CATEGORY

A variable-width character field with a maximum of 100 characters.

Oracle example: VARCHAR2(100).

EVENT_TEXT

A variable-width character field with a maximum of 400 characters.

Oracle example: VARCHAR2(400).

NEWS_TYPE

A variable-width character field with a maximum of 50 characters.

Oracle example: VARCHAR2(50). This field cannot contain a null value.

RP_SOURCE_ID

A fixed-width character field of 6 characters.

Oracle example: CHAR(6). This field cannot contain a null value.

SOURCE_NAME

A variable-width character field with a maximum of 400 characters.

Oracle example: VARCHAR2(400).

CSS – COMPOSITE SENTIMENT SCORE

A numerical field that stores an floating point number value from -1.00 to +1.00 with up to 2 decimal places.

Oracle example: NUMBER(3,2). This field cannot contain a null value.

NIP – NEWS IMPACT PROJECTIONS

A numerical field that stores an floating point number value from -1.00 to +1.00 with up to 2 decimal places.

Oracle example: NUMBER(3,2). This field cannot contain a null value.

PEQ – GLOBAL EQUITIES

A numerical field that stores an integer value of -1, 0 or +1.

Oracle Example: NUMBER(1,0). This field cannot contain a null value.

BEE – EARNINGS EVALUATIONS

A numerical field that stores an integer value of -1, 0 or +1.

Oracle Example: NUMBER(1,0). This field cannot contain a null value.

BMQ – EDITORIALS & COMMENTARY

A numerical field that stores an integer value of -1, 0 or +1.

Oracle Example: NUMBER(1,0). This field cannot contain a null value.

BAM – VENTURE, COMPANY, MERGERS & ACQUISITIONS

A numerical field that stores an integer value of -1, 0 or +1.

Oracle Example: NUMBER(1,0). This field cannot contain a null value.

BCA – REPORTS ON CORPORATE ACTIONS

A numerical field that stores an integer value of -1, 0 or +1.

Oracle Example: NUMBER(1,0). This field cannot contain a null value.

BER – EARNINGS RELEASES

A numerical field that stores an integer value of -1, 0 or +1.

Oracle Example: NUMBER(1,0). This field cannot contain a null value.

ANL_CHG – ANALYST RECOMMENDATIONS & CHANGES

A numerical field that stores an integer value of -1, 0 or +1.

Oracle Example: NUMBER(1,0). This field cannot contain a null value.

MCQ – MULTI CLASSIFIER FOR EQUITIES

A numerical field that stores an integer value of -1, 0 or +1.

Oracle Example: NUMBER(1,0). This field cannot contain a null value.

RP_STORY_EVENT_INDEX

A numerical field that stores an integer value.
Oracle example: NUMBER. This field cannot contain a null value.

RP_STORY_EVENT_COUNT

A numerical field that stores an integer value.
Oracle example: NUMBER. This field cannot contain a null value.

PRODUCT_KEY

A fixed-width character field of 3 characters.
Oracle example: CHAR(3). This field cannot contain a null value.

PROVIDER_ID

A variable-width character field with a maximum of 50 characters.
Oracle example: VARCHAR2(50).

PROVIDER_STORY_ID

A variable-width character field with a maximum of 400 characters.
Oracle example: VARCHAR2(400).

HEADLINE

A variable-width character field with a maximum of 4000 characters.
Oracle example: VARCHAR2(4000).