

ESG news collection and classification methodology

Introduction

The Covalence approach is based on a diversity of sources of information and relies on web monitoring and artificial intelligence together with human analysis.

Stakeholders such as NGOs, governments, trade unions and the media describe the role and activities of companies in positive and negative terms generating either endorsements or controversies.

Since 2001, Covalence has specialized in the semi-automated analysis of such narrative content.

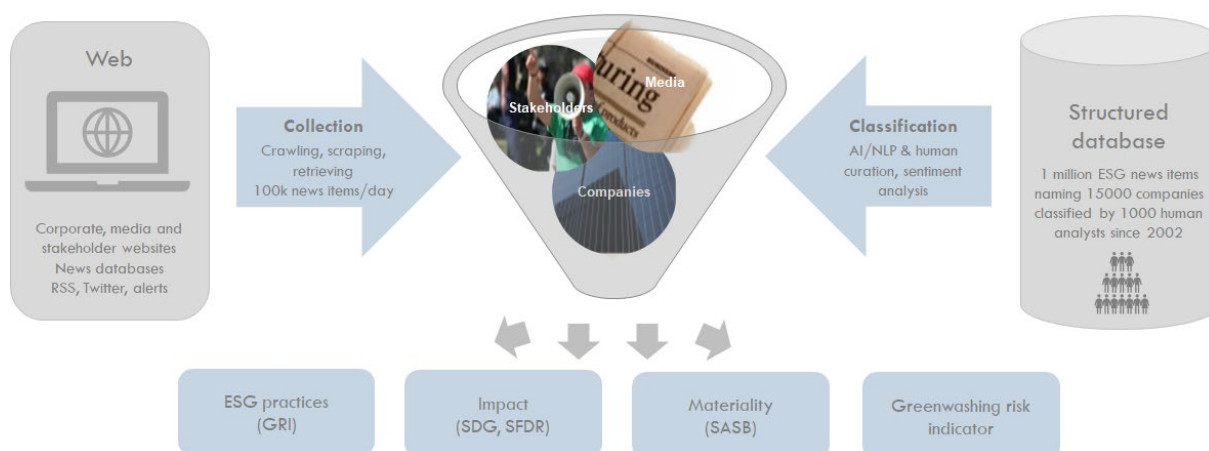
We use data collection and classification tools relying on artificial intelligence techniques (machine learning, natural language processing) in order to analyse the narrative content. This process is reinforced by human interventions to classify the content in terms of polarity (positive/negative) and criteria.

Our team of analysts thoroughly checks entries proposed by the software, thus ensuring high curation standards. Only sources that are publicly identified and available online are considered.

Today, the Covalence database includes more than 12 million documents from over 130'000 different sources on 15000 companies that have been classified and curated by more than 900 analysts in collaboration with over 30 universities.

The database leverages the use of machine learning techniques thanks to the expertise of our Scientific Advisor [Prof. Patrick Ruch](#), field expert and professor at the [University of Applied Sciences and Arts Western Switzerland](#). The use of classification algorithms allows us to fully automate the collection and pre-classification of information including complex information such as polarity – or sentiment – as well as multiple criteria.

Data collection

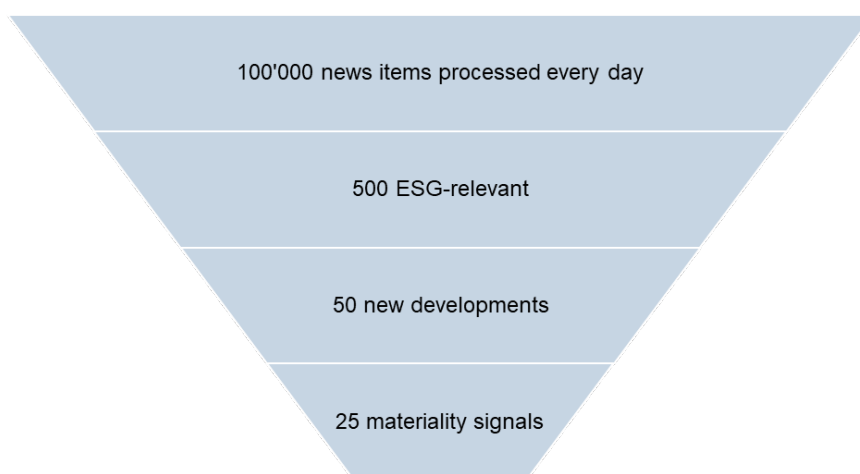


Data classification

Covalence has developed a genuine combination of automated content processing of texts in original language with qualitative assessment by an international team of analysts.

The narrative data gathered by Covalence is made of daily news items gathered from sources such as the media, NGOs, trade unions, etc. Such data is included into Covalence database on a daily basis, and therefore offers opportunities for very dynamic integration of ESG issues into portfolio management.

This data informs on **ESG practices** (GRI-inspired criteria), **impact** (SDGs, SFDR) and **materiality** (SASB), providing signals relevant to quant trading and equity factor investing.



Every day, Covalence processes approx. 100k news items, of which 500 are considered as ESG-relevant. Then, our algorithm identifies new developments, which are more likely to influence markets than mentions of older stories. New developments are flagged when the daily volume of news is abnormal (positive and negative news items are monitored separately). An abnormal volume occurs with a daily volume 3x larger than its 30 days rolling average volume.

The data is also flagged with materiality signals. The idea is to flag ESG news items covering issues that are likely to have an influence on economic and financial variables such as share price. To do so we refer to internationally recognized standards such as the [materiality map](#) produced by the [Sustainability Accounting Standards Board \(SASB\)](#).

Data

Covalence's ESG news database includes 750k+ articles that have been humanly curated and double-checked, in 4 languages: English (70% of total), French (15%), Spanish (10%), and German (5%). This amount grows by 20k per year.

The human classification is undertaken by a team of 4 ESG news analysts employed during their curricular internship in collaboration with various universities in Switzerland and abroad.

Since 2002, Covalence has hosted 900 students operating as ESG news analysts in partnership with more than 30 [universities](#). Over the years we have noticed that a diverse, rotating team of ESG news analysts allows to neutralize the effects of subjectivity biases on classification.

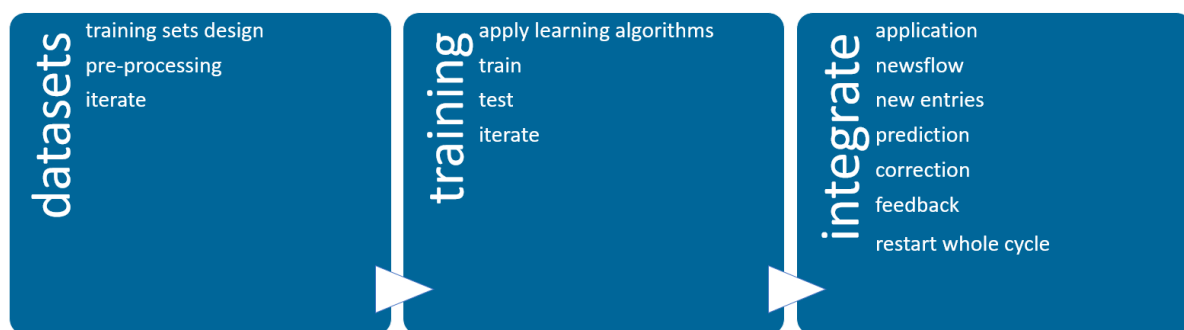
These 750k+ documents are used for training and testing our models for criteria classification (ESG, Global Compact, SDGs, SASB), sentiment analysis, entity recognition, along with meta data linked to each document such as url, URI, country of URI (full list of variables available in annex).

Each variable to be coded needs to be processed with a model. Simple or a little more complex.

Pre-processing

Each classifier is trained on pre-processed data in order to reach its objectives. Here are a few options to consider when pre-processing:

- Stopwords or no stopword;
- Lowercase or not;
- Equally weighted or biased classes;
- List of keywords to include or exclude;
- ...



Every single choice has an impact on the achievement and performance of the resulting model.

The classification models

Criteria, multi-label:

- Supervised multi-label classifiers (logistic regression, maximum entropy) are used for criteria (50) classification: classifiers learn to predict from 0 to five labels for each document;
- Completed with rules-based criteria selection based on proprietary topical dictionary of unequivocal n-grams and combinations (17k and growing entries)

Sentiment, binary:

- Naïve Bayes binary classifier on pre-processed large training sets;
- Forced negative or positive: we force the classifier to choose $>.5 \rightarrow 1$ (positive); $<.5 \rightarrow 0$ (negative)

Entity recognition (companies, countries), heuristics:

- Blend of algorithms, disambiguation rules (Ford \neq John Ford; Total's oil outputs \neq Total oil outputs, etc.)

- Dictionary (company names, brands, subsidiaries, products, false friends, etc.)

Developments groupings

Articles with similar content are grouped using the following unsupervised techniques: clustering, k-means, cosine-similarity matrix, optimization.

This allows for the grouping of new/recent data covering emerging matters/stories.

Each classifier is trained in each of the four languages in use. The whole process is iterative.

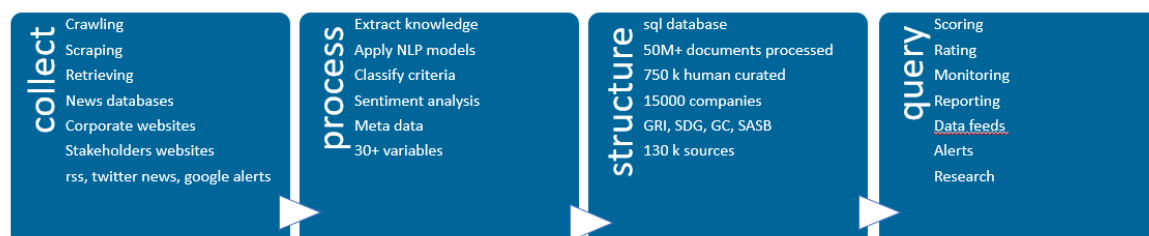
From the lab

We have been reviewing some of the latest developments in NLP such the use of deep learning for text classification, BERT - Bidirectional Encoder Representations from Transformers, FastText or extremeText. We are planning to implement a finetuned version of distilBERT in Q3 2023.

Once the models trained...

Once all the models trained and ready for production, they are integrated into our live applications suite.

Using state of the art extraction techniques and parallel computing we collect and process 100k+ documents a day, from which 1000 documents (similars included) qualify as ESG content for DB inclusion.



Every day human coding and new developments

Considering the high volume of ESG news processed on a daily basis, we prioritize the curation of the most important articles, or “new developments”. Those are identified both in positive news (impact stories) and negative news (controversies).

Event ID 421025 - Carolina Maldonado

Title: Fiscal salvadoreño no descarta que investigación por lavado alcance a Bukele

URL: <https://www.latribuna.hn/2019/10/10/fiscal-salvadoreno-no-descarta-que-investigacion-por-lavado-alcance-a-bukele> Go

Channel: MA-SP Language: Spanish Source: latribuna.hn Domain: www.latribuna.hn

Author:

Description: San Salvador.- El fiscal general de El Salvador, Raúl Melara, no descartó en declaraciones a Efe que una investigación por lavado de dinero contra la sociedad ALBA Petróleos alcance al presidente del país, Nayib Bukele. El pasado 31 de mayo, la Fiscalía General de la República (FGR) allanó la sociedad ALBA Petróleos y otras empresas relacionadas con esta como parte de una investigación que realiza por supuesto lavado de dinero. Una publicación del medio local Factum del pasado 11 de septiembre señala que Bukele supuestamente recibió 1,9 millones de dólares provenientes de la referida empresa en 2013, cuando era alcalde de la pequeña localidad de Nuevo Cuscatlán. (...) Efe preguntó a Melara si este caso alcanza al mandatario salvadoreño y se limitó a señalar que "estas investigaciones se están desarrollando". Bukele señaló el pasado 12 de septiembre, al ser preguntado por un periodista, que en esa época "no era mala palabra hacer negocios con ALBA Petróleos". "Yo no hice negocios con ALBA Petróleos, yo hice negocios con una empresa que hizo negocios con ALBA Petróleos", apuntó el mandatario en una conferencia de prensa. ALBA Petróleos de El Salvador fue formada en 2006 por la estatal Petróleos de Venezuela, S.A. (Pdvs) y alcaldías del Frente Farabundo Martí para la Liberación Nacional (FMLN), ahora en la oposición. El responsable de liderar la millonaria inversión de Pdvs en ALBA Petróleos de El Salvador fue el político salvadoreño y exguerrillero

Highlight: Keywords | Named entities | Disable

Date: 10.10.2019

Orientation: ☐ Positive ☒ Negative

Companies: Alba Se OR ALBA Group OR Interseroh

Criteria List: Social Compliance - 0.87 Fiscal Contributions - 0.55 Corruption - 0.51 Governance

List FR: Select a group... Governance Economic Environment Labor Human Rights Society Product

Countries: El Salvador

Named entities: Fomento, Ministerio Público, Venezuela S, Andorra Norte, ALBA Petróleos, Fiscalía General de, Andorra, BukeleSan Salvador El, Efe, FARC Por, República Gobierno, ALBA Petróleos Bukele, Contratas, Melara Mauricio Funes, Centro América, Gobierno, Douglas Meléndez, Tribuna Una, ALBA Petróleos ALBA Petróleos Salvador, Fiscal, Salvador Raúl Melara, Frente Farabundo Martí, República, Salvador, Nayib Bukele El, Estados Unidos Melara Ministerio Público, Estados Unidos, Panamá Nicaragua Costa Rica, Petróleos, Mauricio Cort, Nuevo Cuscatlán, Nicaragua, Melara, DIARIO, ALBA Petróleos ALBA Petróleos, Liberación Nacional, Estados Unidos Melara, Merino, Bukele, Nota Los, ALBA Petróleos José Luis Merino, ALBA, Mauricio, Honduras, Salvador, El

New developments are flagged when the daily volume of news is abnormal (positive and negative news items are monitored separately). An abnormal volume occurs with a daily volume 3x larger than its 30 days rolling average volume.

Our robot achieves 15-20 cycles a day, thus generating alerts (e-mails sent to our customers) with the same frequency.

Search ESG controversy alerts | Current Folder

All Unread By Date ↑

Today

Covalece SA
eQ3k - ESG controversy alerts
Dear users, 11:00

Covalece SA
eQ3k - ESG controversy alerts
Dear users, 09:58

Covalece SA
eQ3k - ESG controversy alerts
Dear users, 08:46

Covalece SA
eQ3k - ESG controversy alerts
Dear users, 07:23

Covalece SA
eQ3k - ESG controversy alerts
Dear users, 05:06

Covalece SA

eQ3k - ESG controversy alerts

Covalece SA
To

Dear users,

Our alert system has selected the following developments for you:

NEW ESG DEVELOPMENT

2019-10-23

Catalent Inc

Swindon pharma firm given health and safety warning after chemicals incident | <https://www.swindonadvertiser.co.uk/news/17986968.health-safety-executive-gives-catalent-official-warning-incident/>

NEW ESG DEVELOPMENT

2019-10-23

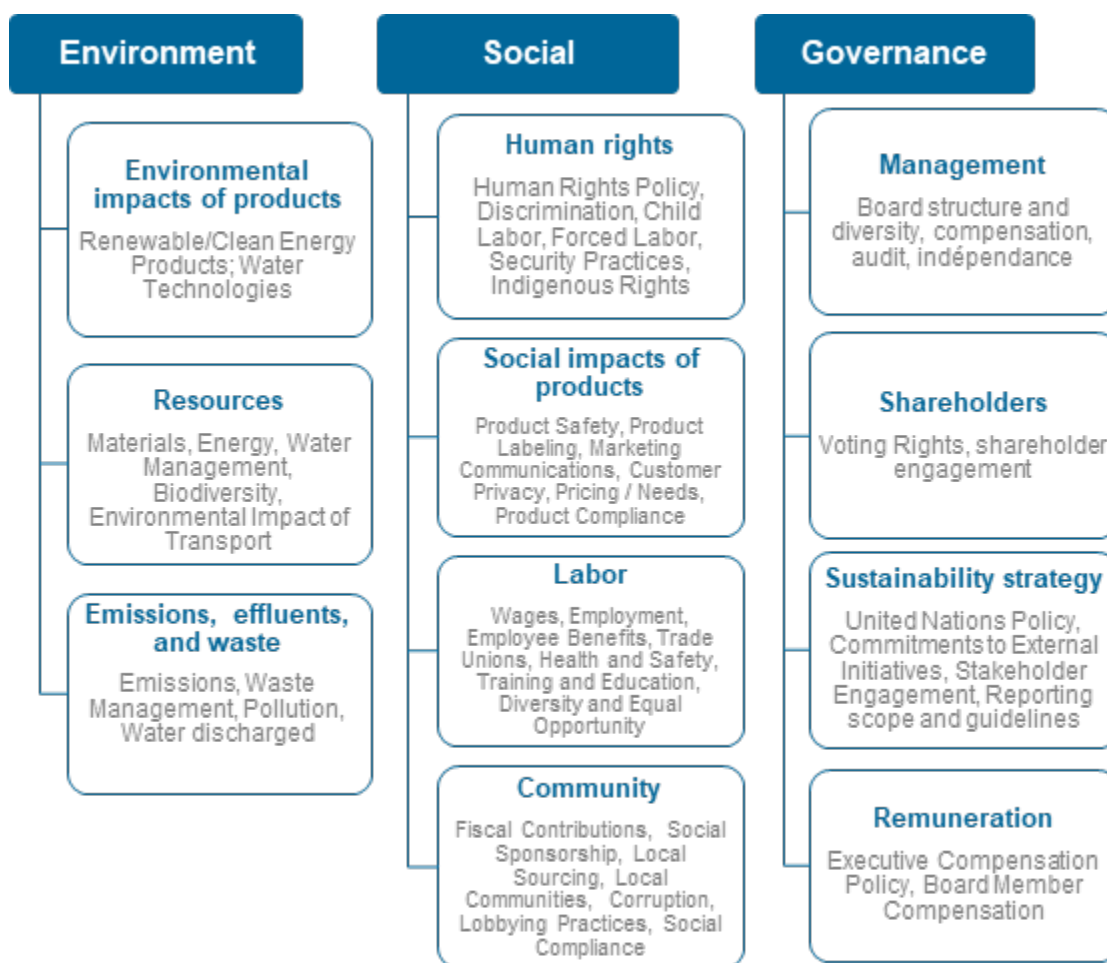
Reply Reply All Forward

Wed 23/10/2019 08:46

Criteria

Covalence first uses a set of 50 criteria inspired by the [Global Reporting Initiative's](#) sustainability reporting guidelines. These criteria serve to classify the narrative content which is gathered thanks to our semi-automated search process using a broad set of sources.

The data is then recoded with hundreds of topics and sub-topics and organized into 11 dimensions within 3 categories: Environment, Social, Governance.



SDG mapping

The data is also classified with the 17 Sustainable Development Goals to show companies' mapping to the SDGs to provide insightful material for impact analysis and to support thematic investment strategies.

Sources

Covalence gathers online information using media monitoring platforms and scraping individual websites. Only identified information sources are used, we don't consider anonymous comments.

Media monitoring platforms

Media monitoring platforms are the main providers of news aggregated by Covalence. They are used to gather information from hundred thousand potential sources.

Individual sources

Covalence also directly monitors individual websites that regularly publish relevant content, using in house web crawling, scraping, and retrieving capacities. Crucial is the diversity of sources used among media, NGOs, trade unions, international organisations, governments and academia.

Languages

Information is searched for in four languages: English, French, German, Spanish.

Neutrality

Covalence does not see some sources as more reliable than others. Any source is considered equally. Covalence does not validate information sources, neither the content of information. What we do is collect, confront and synthesize the maximum of relevant documents from different sources. Our policy is to put ourselves in the position of an independent newspaper in front of statements, opinions, reader's letters: publish any information provided it has relevance and an identified author, without endorsing its content.

Equal weighing of individual sources

Covalence follows a principle of equal weighing of individual sources, relying on the spirit of democratic debate and consensus. The "size" of source (audience, quantity of readers / viewers) is not taken as a weighting criterion, neither is placement in print press. Following are our arguments for applying such an equal weighing approach:

a) The modern world is characterized by social complexity, cultural diversity, ethical pluralism and scientific uncertainty: considering "small" sources at the same level as "large" ones is a way to cope with such complexity and diversity.

b) It is technically difficult to measure the size, or popularity of sources and find a weighing factor for such an heterogeneous ensemble of sources as large medias, specialized NGOs, individual correspondents and multinational companies' headquarters.

c) Western and Anglo-Saxon sources are overrepresented in Covalence database, because such sources are more numerous online and are more easily accessible than others. Applying a weighing factor could amplify the already existing overrepresentation of Western and Anglo-Saxon sources.

d) Some search engines email alerts used by Covalence only cover pages with the highest popularity (page rank): for a part the most popular pages are already naturally selected.

e) Echoes, repetition make weigh. If an obscure blog publishes a story, it might enter our system but if this story is not picked up by other sources its weight will be marginal. On the contrary, if a credible journal publishes a story, it is likely that other sources will repeat it. This will produce several points in our system and will have an effect on a company's ESG reputation score. This is how a weighing process is naturally working: the system measures the noise made by news, the echoes generated by a story among numerous sources. Rather than one particular document, it is the aggregation of a large number of documents that gives a significant picture of reality.