



Extraggo®

multilingual text analytics
to get the gist of any text

With Extraggo® you can distill and organize knowledge from a text. Thanks to its seamless combination of statistical and semantic techniques, **Extraggo®** is the system that allows you to **get meaningful insights out of text**, independently of the language used in the input data. It can:

- **Recognize languages**
- Extract and rank **keywords, concepts, entities** and **domains** involved in a text and **interconnect** them in an intelligent way
- **Generalize terms to concepts and named entities**, independently of the language
- Detect **cross-lingual semantic similarity**
- Represent key concepts in a **knowledge graph**

Terms, concepts and entities are **ranked by importance**, and the domain information conveyed in the text is provided. Its special asset is the **high precision of domain extraction and classification**, with more than 600 domains covering every field of human knowledge, ranging from art to science, politics, sport, law, computing, finance, and much more.



LANGUAGES SUPPORTED

Extraggo **fully supports 10 languages.**

AR

Arabic

ZH

Chinese

NL

Dutch

EN

English

FR

French

DE

German

IT

Italian

PL

Polish

PT

Portuguese

ES

Spanish

More languages are available upon request.

Uses of Extraggo®

SOME EXAMPLES

DOMAIN EXTRACTION



Given a data collection, Extraggo® mines **keywords** and **domains** and maps them by topic.

This is fundamental for any text analysis task, but it also makes it easy to get the gist of a text. Turning Extraggo® into a crucial tool in many fields, including **cybersecurity, news monitoring** (monolingual and multilingual), **disaster prevention, public health, governmental companies** and **organizations**.

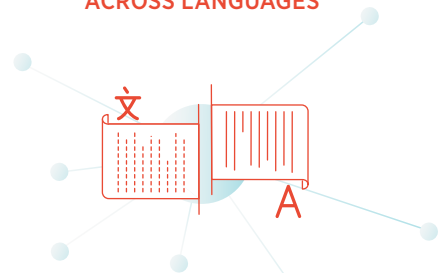
KNOWLEDGE MANAGEMENT



The Extraggo® system analyzes data and organizes them by topic. A **smart organization** of data is the first step to making them more useful and **optimizing knowledge management**.

Whereas traditional search was based on keywords, semantic search is based on **concepts** and **search intent**. This is critical for **publishing houses, press companies, and newspapers** in order to optimize and **personalize results** and to suggest **related contents** to their readers.

DOCUMENT COMPARISON ACROSS LANGUAGES



Extraggo® distills knowledge from text written in **multiple languages**.

Furthermore, concepts and entities are linked to **WordAtlas®, our multilingual knowledge graph**, which enables language independence.

This makes it possible to compare the core of documents **independently of the input language**, facilitating **academic or scientific research** in any field.

How it works

GIVEN A TEXT, EXTRAGGO PROCESSES IT AND EXTRACTS THE INFORMATION

DOMAINS | NATURAL MATERIALS ● | EXPLORATION ● | EVOLUTION ● | SCIENTIFIC BUILDINGS ● | NUTRITION ●



Mars could be teeming with glowing green microbes, according to researchers who found the bacteria living in volcanic rocks deep beneath the ocean floor.

A wide variety of bacteria have been found thriving in rocks deep below the Earth's oceans in densely packed communities and the same is likely true of Mars.

Researchers from the University of Tokyo discovered the single-celled organisms after spending a decade studying rocks found in the South Pacific Ocean in 2010.

Last year NASA's rover Curiosity identified clay-rich regions on the Red Planet and the Tokyo team expect it to be similar to the minerals holding the green bacteria.

NAMED-ENTITIES

- Mars
- South Pacific Ocean
- Tokyo
- NASA

CONCEPTS

- bacteria
- mineral
- ocean floor

DOMAIN IDENTIFICATION

Given a text, Extraggo® identifies and ranks the domains (i.e. the fields of knowledge) of the text, such as health, government, industries, crime, cultures, economics and hundreds more, making text classification and organization easier, also across languages.

NATURAL MATERIALS

GREEN MICROBES BACTERIA VOLCANIC ROCKS ...

LANGUAGE DETECTION

Extraggo® features high-performance language detection on the input text, which increases the quality of the concept and entity extraction and classification.

ENGLISH - DETECTED



WORD SENSE DISAMBIGUATION

By determining the overall meaning of the entire input text, Extraggo® successfully disambiguates ambiguous words and terms by associating them with the most suitable meanings in context linked to WordAtlas®, our large-scale multilingual knowledge graph.

...similar to the minerals holding the green bacteria.



Solid homogeneous inorganic substances occurring in nature having a definite chemical composition

NAMED ENTITY RECOGNITION

Extraggo® classifies Named Entities (i.e. real-world entities that can be denoted with a proper name) into categories such as person names, locations, organizations, etc. Categories can be personalized.



Pacific Ocean
Mars/Red Planet
Tokyo



NASA

CONCEPT EXTRACTION

Extraggo® identifies the key concepts in the text: as opposed to named entities, a concept is a general, abstract notion, like virus, conviction, bailout, etc., including domain-specific concepts like deep learning, cash flow statement, airspeed indicator, etc. As a result, Extraggo® summarizes the text into a cluster of language-independent meanings.



bacteria, mineral, ocean floor, single-celled organisms...

KNOWLEDGE GRAPH CREATION

Given a text, Extraggo® returns a visual, structural representation of concepts and entities and their mutual relations by constructing a document-specific knowledge graph.

Thanks to our innovative myKnowledgeGraph service, you can connect several such graphs constructed from a range of your documents to produce an aggregated enterprise graph.



NASA

BACTERIA

MINERAL

MARS

A glimpse inside the machinery

EXTRAGGO® IS INTEGRATED WITH OUR COMPANION PRODUCTS
COMPREHENDO®, WORDATLAS® AND NLP PIPELINE



COMPREHENDO®

Comprehendo®, the core engine of Babelscape's multilingual text understanding, identifies the proper meaning of a dictionary word in a given context. It works in tandem with **Extraggo®**, extrapolating the meaning of the text and **linking ambiguous words** explicitly to **concepts** and **entities**.

This system works with both **standard text** and **text snippets** and brings two main advantages:

- It enables the **semantic aggregation** of similar information written using different words, either in the same language, or in different languages
- It enables **discrimination between meanings** of the same word



WORDATLAS®

In **Extraggo®**, concepts and entities are linked to **WordAtlas®**, our next-generation **multilingual knowledge graph**. WordAtlas® provides **millions of lexicalizations for each language**, from common nouns, adjectives, verbs and adverbs, to hundreds of thousands of technical terms and millions of named entities, such as people, locations, organizations and products.

Each meaning comes with a wealth of information, including:

- **synonyms** in dozens of languages
- **definitions** (often multiple) in each language
- **taxonomical information:** generalizations and specializations
- thousands of **different relations**



NLP PIPELINE

Extraggo® is a module of our **Natural Language Processing Pipeline**, which performs tasks like language recognition, morphological analysis, part-of-speech tagging, lemmatizing and named entity recognition.

In addition, our pipeline also features:

- **large scale**, making it possible to process millions of texts in seconds
- **parallelism of independent modules**, which can run dozens of tasks independently
- availability both as an **online** service and as an **offline** software package

We have developed our pipeline giving priority to the highest levels of **vertical** and **horizontal scalability**. Our pipeline is ready to be applied on **cloud infrastructures** through the use of using the most common orchestrators.

Pricing

Extraggo has been conceived with a **modular pricing model**, so as to be a **flexible solution** to your specific business requirements.

You can choose the number of languages, go for online vs. offline usage and ask for customized domains. Every company is unique, and that's why we will create a **customized plan** for you.

Contact us at info@babelscape.com.



VARIABLE NUMBER
OF LANGUAGES



ONLINE AND OFFLINE
VERSIONS AVAILABLE

CUSTOMIZABLE
DOMAINS

