WordAtlas®

the next-generation **multilingual knowledge graph** based on BabelNet

Knowledge graphs are the 21st century counterpart of dictionaries and encyclopedias in previous centuries. They **organize knowledge into a coherent network of meanings** and they **enable Artificial Intelligence applications w**hich exploit this knowledge to perform text understanding.

WordAtlas® is a key resource for **multilingual Natural Language Understanding**. What makes WordAtlas® special is its **linkage between concepts and words in hundreds of languages**: 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.

WordAtlas® comes with a wealth of information, including:

- **synonyms** and **multiple definitions** in dozens of languages
- taxonomical information: **generalizations** and **specializations**
- **280 million** different lexico-semantic relations
- **pictures** which illustrate concepts
- a **flexible**, **computational** organization of **multilingual knowledge**

OUR CUSTOMERS USE WORDATLAS FOR:

SEMANTIC SEARCH

obtaining high accuracy of results through an excellent understanding of data in their contextual meaning

MULTILINGUAL CONTENT RETRIEVAL

having access to relevant semantic-based content, overcoming language barriers and saving time

SCALING DATA TO MANY LANGUAGES

making data or concepts available in one or more target languages

KNOWLEDGE GRAPH CREATION

creating a semantic network based on hundreds of millions of concepts and their mutual relations

Anatomy of a multilingual synset

A **synset** is a set of senses that represent a given concept in **multiple languages**. A synset includes all of the **synonyms** for each language that are **interchangeable in a certain context**.

Total number of synsets

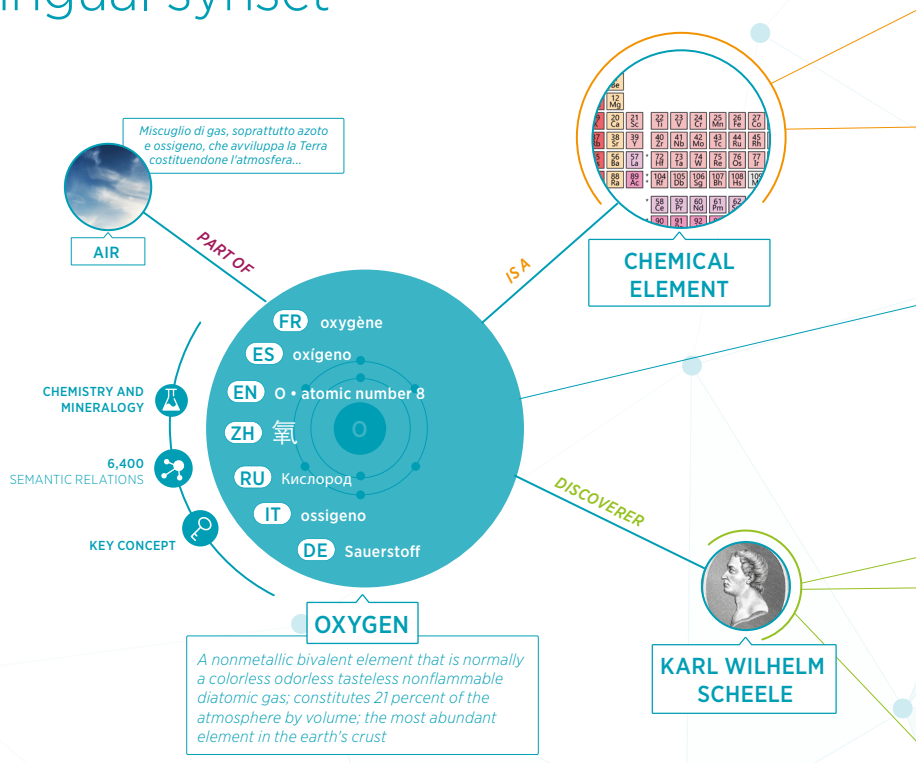
16 million

Synsets with pictures

11 million

Synsets with domain tags

3 million



Babelscape

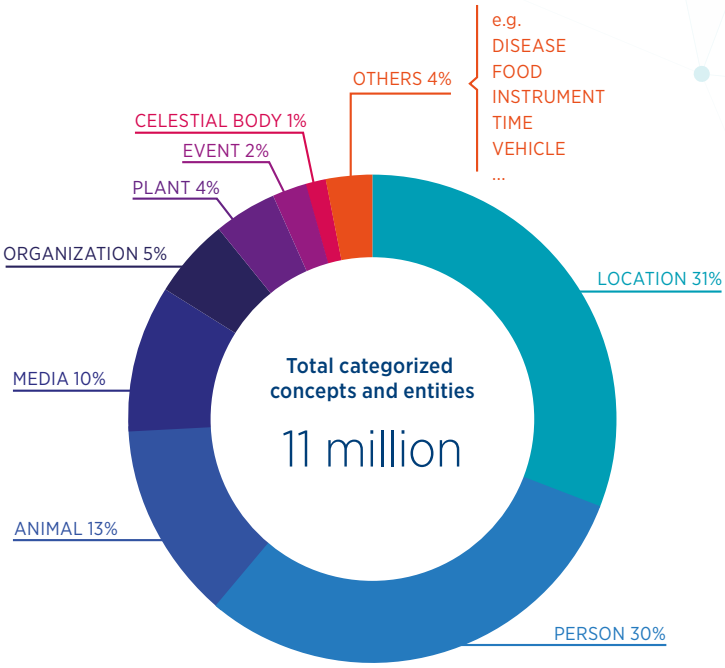
http://babelscape.com | Email: info@babelscape.com

Statistics about WordAtlas

16 million

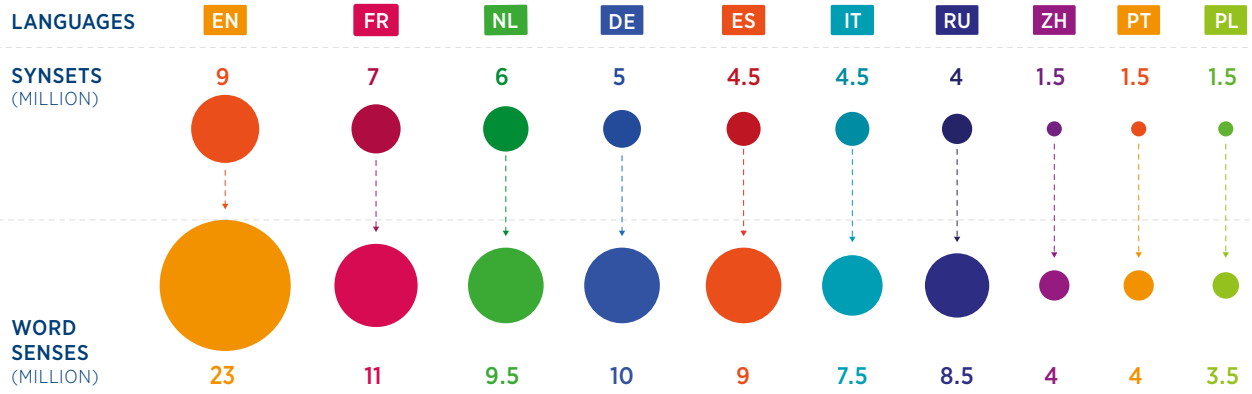
CONCEPTS AND
NAMED ENTITIES
IN WORDATLAS®

65% categorized



284

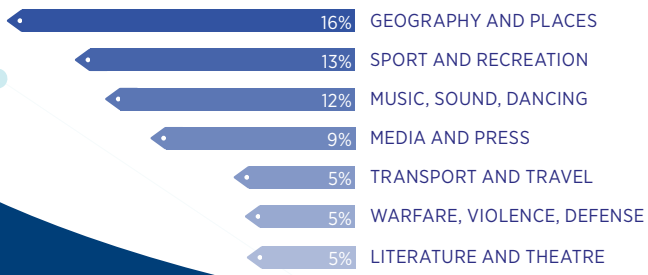
SUPPORTED
LANGUAGES
INCLUDING:



42

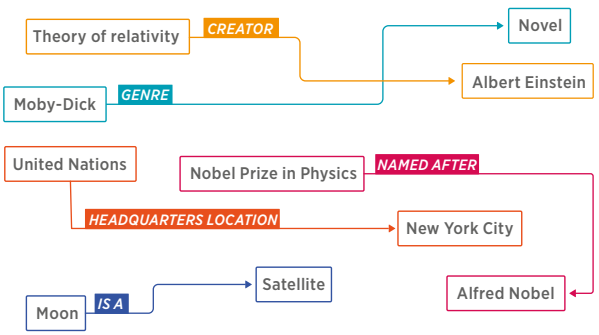
DOMAIN TAGS

More than 2.8 million WordAtlas® synsets are **labeled with domains**. The most common are:



2.5 billion

SEMANTIC RELATIONS



Babelscape

http://babelscape.com | Email: info@babelscape.com

3 Editions

EACH ONE COMES WITH
MANUAL ANNOTATION OF

SILVER EDITION

- Curated **mapping** between lexical-semantic resources
- **Geolocalization** of concepts and entities

GOLD EDITION

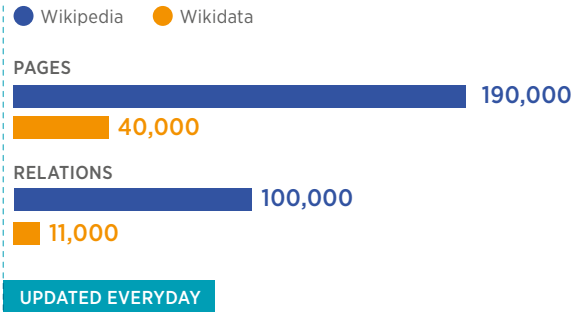
- **Domain tags**
- **Is-a** information
- **Regional and dialect information** of language use

PLATINUM EDITION

- **Named Entity** vs. **concept** tags
- 18 Named Entity **categories**

Live

UPDATES



Every day WordAtlas® ingests **new data** coming from open sources, such as Wikipedia and Wikidata.

More **quality content** means wider multilingual **coverage** and better **language understanding.**

EXTRA FEATURES

- **Professional translations** of thousands of synsets into tens of languages
- Fine-grained and coarse-grained **sense distinctions**
- A **fundamental dictionary** of 20,000 synsets
- **Probability distributions** of word senses
- **Offensive**, **netspeak**, **prefixes**, etc.

READY FOR USE IN
DEEP LEARNING ARCHITECTURES:

- **Latent** and **explicit vector representations** of meanings
- Sense-annotated **corpora** in multiple languages

Synset and sense

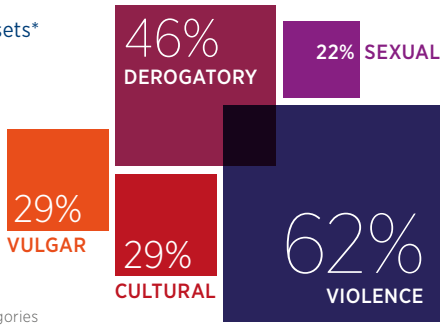
CLASSIFICATION

We manually classified concepts and word senses based on a number of criteria: **offensiveness** and **cultural bias**, informal use on the Internet (**netspeak**), **prefixes**.

Total **OFFENSIVE** synsets*

6 thousand

including:



*every synset can be included in one or more offensive categories

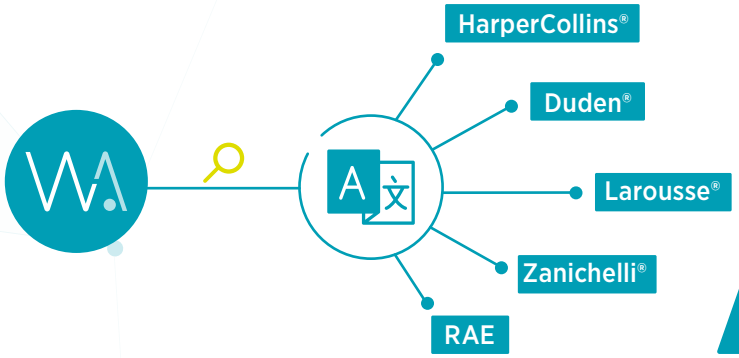
LANGUAGES	EN	FR	DE	ES	IT	NL	PT	PL
NETSPEAK								
SYNSETS	370	190	90	190	250	210	120	130
SENSES	900	400	160	260	400	320	125	170
PREFIX								
SYNSETS	1,450	1,160	660	1,120	1,200	610	1,450	1,460
SENSES	2,400	1,550	900	1,400	1,790	380	2,100	1,840

Babelscape

http://babelscape.com | Email: info@babelscape.com

Linked

TO THE MOST
REPUTABLE DICTIONARIES



READY TO BE CONNECTED

to large multilingual resources like **UMLS®** - the Unified Medical Language System - so as to enable tasks such as **information extraction** and **multilingual semantic search**.

Linkage available also to **EuroVoc**, **IATE** and other **domain glossaries**.

Knowledge graph APIs

WordAtlas® comes with **high-performance APIs** for **Python** and **Java** (therefore supporting all JVM-based languages, such as Kotlin, Scala and Groovy). The API enables access to the multilingual knowledge graph and includes a wide range of methods for:

- searching by **word** and **concept ID**
- retrieving information about concepts and entities, such as **multilingual lexicalizations**, definitions, examples, semantic relations, links to images and much more
- supporting **functional programming** with lambda functions for querying and retrieval purposes.

Plans

WordAtlas® is available both **online** and **offline**.

Discover the most suitable solution for your business. Contact us at info@babelscape.com and we will formulate a **customized plan**.

	SILVER	GOLD	PLATINUM
Python API	✓	✓	✓
Assistance	-	✓	✓
Customizability	-	-	✓

Babelscape

http://babelscape.com | Email: info@babelscape.com