KNOWLEDGE GRAPH

## The Brains Behind Symbolic AI

airplane

The Fundamentals of the Knowledge Graph

The expert.ai knowledge graph is a lexical database structured in a network framework. This extensive proprietary language ontology was developed over many years and is constantly updated.

Unlike traditional dictionaries where words are listed in alphabetical order, words contained in this database are arranged in nodes that contain groups of items expressing identical or similar meaning.

These groups are connected to each other in a logical language-based format via semantic links.
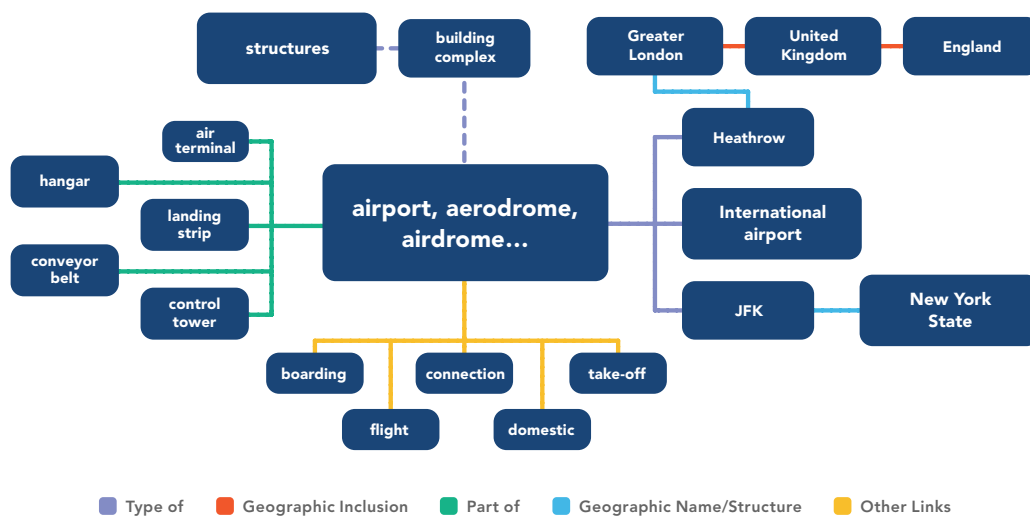
In the expert.ai knowledge graph, "words" are called **lemmas** and groups of lemmas are called **syncons**. The relationships between syncons are called **links.**

As every language has its own lexicon and rules, different knowledge graphs have been created, one for each language managed by expert.ai's tools and platforms.

All examples presented in the following pages are taken from the knowledge graph for the English language.

These particular groups of tokens include:

- 430,000+ syncons;
- 510,000+ lemmas;
- 80+ relation types;
- rules for inflections.

The extended English knowledge graph integrates knowledge and entities from Wikipedia and contains more than 6 million syncons and more than 151 million links.

However, the definitions are valid for any knowledge graph, because they refer to the general architecture of the semantic network shared across different languages.

Syncons and Lemmas

Each syncon coincides with a node on the knowledge graph and represents a concept in a language. Each syncon is made up of:

- one or several lemmas that express identical or similar meaning;
- a set of attributes making explicit, in a machine-readable format, all grammatical, syntactical and semantic characteristics belonging to lemmas and concepts;
- a univocal identifier, called **syncon ID**, which is a whole number made of one or several digits.

Syncons

The term syncon blends the Latin words **synonimus congressus**, meaning "group of synonyms." Every syncon is designed to represent a different **concept** in a language based on the principle of **synonymy** (from the Greek **syn + nymy**, meaning "same name").

This principle states that different signifiers, or the form a concept can take in language (e.g., lemmas) can refer to the same signifier (the concept itself) and, thus, convey the same meaning. In fact, every syncon contains all synonymous lemmas used by native speakers to express the same concept.

plane — fixed-wing aircraft

aeroplane — airplane

In essence, the synonyms of a syncon all have the same meaning. Therefore, a concept in a knowledge graph is represented by a set of synonyms. For example, the concept represented by the image above, defined as "a powered flying vehicle with fixed wings," can be expressed in English using the words: "aeroplane", "plane", "airplane" and "fixed-wing aircraft".

There are two ways in which one can speak of synonymy. The first, known as **strict** or **perfect synonymy**, refers to a relationship between two or more words that can be used interchangeably in all sentence contexts. The second, called **loose synonymy**, refers to a relationship between two or more words that are often, but not always, used interchangeably.

Strict synonymy is rare, as it is difficult to find two perfect synonyms. Specialty languages offer some good examples (e.g., "antipyretic" and "febrifuge" are perfect synonyms), **but knowledge graphs aim to represent the common lexicon of a language**, which is much more likely to express meaning in many different shades with less clear-cut term choices. For this reason, syncons have deliberately been created to group loose synonyms representing the same concept.

Two other linguistic principles are involved in knowledge graph design: **polysemy** and **homonymy**.

Polysemy

Polysemy is the capacity for a word to have multiple related meanings. For example, "bank" can refer to a financial institution as well as the building used by such an institution. They have the same etymology and, therefore, refer to the same original idea.

Homonymy

Homonymy describes the case in which the different meanings expressed by a word are not connected or related. For example, "bank" could refer to a slope or a long-raised mass of earth or it could refer to a financial institution. There is no relation between the two terms.
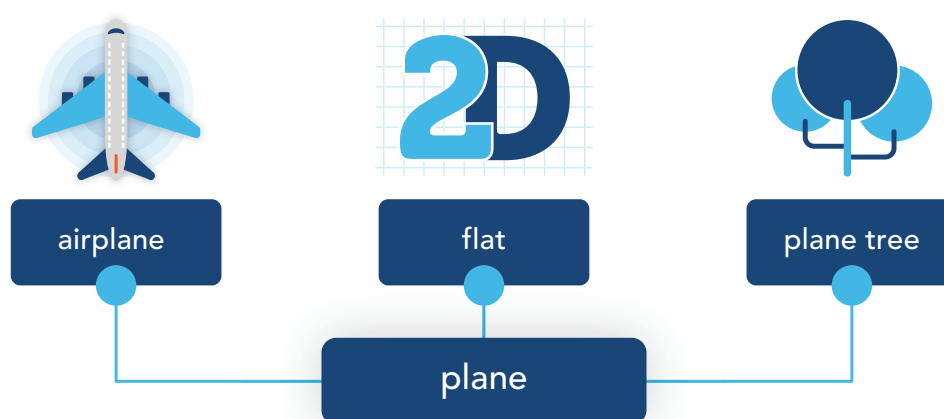
Traditional dictionaries list a single entry for each word without distinction between monosemic and polysemic words (or homonyms). A knowledge graph contains a syncon for each meaning of a word. Therefore, a single lemma can be contained in several syncons.

For example, the lemma "plane" can express more than one meaning and is therefore part of more than one syncon. It could refer to "an airplane", "an infinite two-dimensional surface" or "a level or flat surface".

Lemmas

Lemmas are a fundamental component of a syncon and a key access to concepts. **A lemma is the base form of a word representing all its inflected forms.** The base form of each lemma depends on the part of speech it belongs to, such as:

- the singular for a noun ("child" also represents "children");
- the bare infinitive for verbs ("go" also represents "went", "goes", "going");
- the positive form for adjectives and adverbs (which also represent their comparative and superlative forms).

There is no need to add all inflected forms to a knowledge graph since, during text disambiguation, these are recognized and associated with their base form in the knowledge graph.

Lemmas can be found in the knowledge graph in both lower and upper case, depending on their proper spelling. For example, in the English language, nouns are usually found in lower case, but the names of languages, months and days of the week are written in upper-case.

Adjectives are also generally in lower case, but not for adjectives related to nationality. However, proper nouns such as cities, companies and organizations are always found with an upper-case letter.

In a knowledge graph, a lower-case lemma will also be recognized in a text if it is found with one or more upper case letters. By contrast, a lemma with one or more upper-case letters will not generally be recognized in a text if it is found with a different spelling than the lemma in the knowledge graph.

However, we have integrated this general functionality with some tools that enable the engine to consider certain situations (e.g., social media content in which the correct spelling is not always followed).

Lemmas can be any of the following:

- Single headwords ("plane", "cat", "eat", "furious", "eagerly")
- Compound words ("white-collar", "dog catcher", "check in")

- Collocations (i.e., elements that often co-occur in a language denoting a specific concept that is different from the concept represented by the single elements considered in isolation; "degree of purity", "earth movements", "air corridor")

- Abbreviations or acronyms ("sync", "B.L.A.", "AmEX")

- Idiomatic expressions (i.e., combinations of words that have a figurative meaning; "be in deep water", "cash cow", "in the blink of an eye")

In general, a knowledge graph contains **lexicalized concepts**, which are concepts that can be represented by a single word, a collocation or an idiomatic expression. However, a well-defined concept does not always correspond to a specific lexical item and may require a description to be conveyed.

Sometimes, for organizational and consistency reasons, it is necessary to fill in these **lexical gaps** by adding phrase to knowledge graph such as "generic animal activity", "vegetal feature", "positive human non-physical feature".

These lemmas belong to the many "functional syncons" at the highest and most abstract level of knowledge graph hierarchy; these particular syncons and lemmas will be fully described in a dedicated article.

As syncons, each lemma is provided with a set of attributes making explicit, in a machine-readable format all grammatical, syntactical and semantic characteristics belonging to lemmas.

Customizing and Extending Knowledge Graph

The knowledge graph is not meant to be an encyclopedic reference tool for human users, but a core module engineered to interact with the semantic disambiguator to achieve text disambiguation.

Indeed, it provides a machine-oriented representation of a language's lexicon where every item is provided with a set of attributes making explicit, in a machine-readable format, all grammatical, syntactical and semantic characteristics belonging to words and concepts.

The knowledge graph is the heart of expert.ai's core technology for natural language processing (NLP), which is currently being used by major customers across vertical markets (government, security and intelligence, financial services, insurance, publishing, oil and gas, etc.).

With each new expert.ai Platform release, the knowledge graph is enriched, improved and refined to better address a variety of applications and use cases.

In addition, the knowledge graph can be customized to specific domains to include domain-specific terminology and relations and improve the quality of results.

The expert.ai Platform allows customers, partners and professional services teams to easily extend and customize knowledge graph with syncons and lemmas related to their specific business terminology, and to their product and service offerings. In addition, it provides machine learning-based functionalities to augment the knowledge graph automatically or semiautomatically.

Knowledge is everything when it comes to understanding language. So make the commonsense decision to integrate the knowledge graph into your language models today.

Get Started

Interested in learning how NLU AI will transform your company? Get started here.

See what expert.ai can do for you!

expert.ai

About us

Expert.ai is the premier artificial intelligence platform for language understanding that augments business operations, scales data science capabilities, simplifies AI adoption and provides the insight required to improve decision making throughout organizations. The expert.ai brand is owned by Expert System (EXSY:MIL), that has cemented itself at the forefront of AI-based natural language solutions across Insurance, Banking, Publishing, Defence & Intelligence, Life Science & Pharma, Oil Gas & Energy, and more.

🌐 **www.expert.ai**   💬 **info@expert.ai**