**Semantics**is a branch of linguistics studying the meanings of words, their symbolic use, also including their multiple meanings.

“One morning I shot an elephant in my pajamas. How he got into my pajamas I’ll never know.” Groucho Marx

This sentence is semantically ambiguous: it’s not clear if the author shoots an elephant while in his pajamas, or if he shot the elephant, which happened to be in his pajamas. Clearly, in this example only one of the two makes sense, but both cases are gramatically correct.

*“John and Mary are married.” (To each other? or separately?)*

*“John kissed his wife, and so did Sam”. (Sam kissed John’s wife or his own?)*

[More information on linguistic ambiguity](https://cs.nyu.edu/faculty/davise/ai/ambiguity.html).

**Lexical search engines**

At first, **search engines**(Google, Bing, DuckDuckGo, Yahoo, etc.) were *lexical:*the search engine looked for **literal**matches of the query words, without understanding of the query’s meaning and only returning links that contained the exact query.

For example, if the user looked for “cis lmu”, the homepage of the CIS centre from the LMU university matches the query, since:

* the CIS centre’s homepage contains these 2 words
* the url of the homepage contains these 2 words
* the page is at the top level of the domain
* and many other reasons specified by the search engine

All these criteria are easy to check, and they alone make this page a very good candidate for the top hit of this query. No deeper understanding of what the query actually “meant” or what the homepage is actually “about” were needed.

**Semantic search**

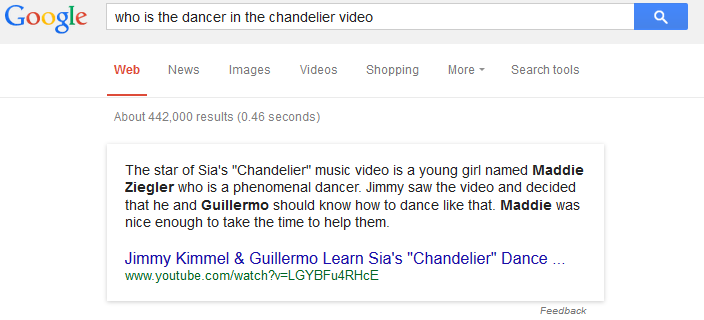
[Semantic search](https://en.wikipedia.org/wiki/Semantic_search) is **search with meaning**. This “*meaning*” can refer to various parts of the search process:

* understanding the query, instead of simply finding literal matches,
* or representing knowledge in a way suitable for meaningful retrieval.

Semantic search goes beyond the ‘static’ dictionary meaning of a query to understand the searcher’s intent within a specific context. By learning from past results and creating links between entities, a search engine can make use of contextual meaning of terms as they appear in the searchable database to generate more relevant results.

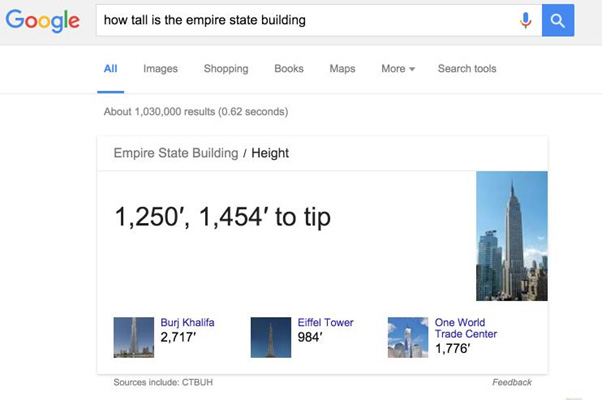
It also allows the user to ask **natural language questions**, as opposed to adapting our language for computers to understand: ‘how do I start a career in data science?’ vs. ‘data science career steps and tips’. In the second case, there are no verbs or unecessary words, just keywords that the user believes are relevant to the search engine.

Semantic search results also require that information from several different sources is brought together to answer the query satisfactorily.



In this example, the main result is a YouTube video about Jimmy Kimmel and Guillermo regarding Maddie Zieger, the ‘star of Sia’s Chandelier music video’.

* Google ‘understands’ that the query ‘who is X’ must have a person’s name as a result.
* Note that both ‘Maddie Ziegler’ and ‘Guillermo’ are highlighted, and this is an incorrect result from Google. On the other side, ‘Jimmy’ is not highlighted. Probably because Guillermo is closer to the verb ‘dance’ in the sentence than ‘Jimmy’ is. For more advanced readers, you might notice that the pronoun ‘he’ in the third line refers to Jimmy, and both men are in the same category and therefore equally close to the verb ‘dance’, but successfully linking ‘he’ to ‘Jimmy’ is another linguistic problem, called *coreference resolution*and is not resolved well in this example. ([Wikipedia link,](https://en.wikipedia.org/wiki/Coreference#Coreference_resolution) [Stanford NLP Group’s implementation](https://nlp.stanford.edu/projects/coref.shtml))
* There is no literal match for ‘dancer in chandelier video’ with ‘star of chandelier music video … who is a phenomenal dancer’. The words don’t appear next to each other, yet the search engine makes the connection between ‘star in a music video’ and ‘dancer’.



In this example, the result is not only correctly displayed but it has a user-friendly section with a picture, as well as other similar buildings and their heights.



The examples displayed are Google results. Even if other search engines have imlpemented semantic search functionality in recent years, Google was the first to do so, with the [Hummingbird update in 2013](https://www.searchenginejournal.com/google-algorithm-history/hummingbird-update/#close), and the most accurate one as to date.

**Bonus:**[**how does it work?**](https://medium.com/value-stream-design/semantic-search-2371fda0bbf5)

Google included a Knowledge Graph in 2012, an [ontology,](https://en.wikipedia.org/wiki/Ontology_%28information_science%29) a representation of semantic relations between people, places and things in a graph format. These relations can be synonyms, homonyms, etc. With the Hummingbird update in 2013, Google had a huge knowledge graph of its collection of around [570 million concepts and relationships](https://www.crazyegg.com/blog/google-hummingbird-change-approach-seo/).

When a new query arrives to the system, first of all the query is broken into root terms, by using natural language processing (NLP) algorithms such as POS tagging retrieval, named-entity recognition, error correction, conversion to word embeddings, search for synonyms, etc.

Afterwards, these terms are matched into the ontology, obtaining the closest terms from the huge graph. These terms or links are the ones more relevant to the input. Good systems make the ontology independent from language, such that a query in Spanish can be matched to ontology terms in English.

**Other links**

[1]: Bast, Buchhold, Haussmann, [Semantic Search on Text and Knowledge Bases](http://ad-publications.informatik.uni-freiburg.de/FNTIR_semanticsearch_BBH_2016.pdf), University of Freiburg, 2016

[2]: Mangold, [A survey and classification of semantic search  
approaches,](https://www.semanticscholar.org/paper/A-survey-and-classification-of-semantic-search-Mangold/344885b612ddbe5ede930571bd433f195245147d) University of Stuttgart, 2007

[3]: Fatima, Luca, Wilson, [New Framework for Semantic Search Engine](https://www.researchgate.net/publication/262635559_New_Framework_for_Semantic_Search_Engine), Anglia Ruskin University, 2014

Github Engineering, [Towards Natural Language Semantic Code Search,](https://github.blog/2018-09-18-towards-natural-language-semantic-code-search/) 2018

Hamel Husain, [How To Create Natural Language Semantic Search For Arbitrary Objects With Deep Learning](https://towardsdatascience.com/semantic-code-search-3cd6d244a39c), 2018