

KeyQ

The KeyQ system makes it possible to extract the simple and compound terms (i.e. the terminology of the field of application) from a body of texts. It also allows users to search for those terms, returning the paragraphs where they are found in order of relevance.



Contact information

Address: ETSI Informáticos - UPM
Ontology Engineering Group (OEG)
Departamento de Inteligencia Artificial
Universidad Politécnica de Madrid
Campus de Montegancedo, sn
Boadilla del Monte, 28660 Madrid
Phone number: 910672915
Website: demo.inno.oeg-upm.net
Email: mariano.rico@upm.es

Technological Offers type

Technological solutions

Research and innovation areas

- Digital Technologies, Artificial Intelligence, Cybersecurity, 5G, Robotics

ODS



Available from: 2019

Where?

Artificial Intelligence R&I Centre (AI.nnovation Space) Ontology Engineering Group

Files

[Download additional documentation \(pdf\)](#)

Keywords: | [information](#) | [retrieval](#) | [smart search](#) | [terminology](#) | [terms](#)

Description of the technological basis

KeyQ is a system that extracts the 'compound terms' (i.e. terminology of the field of application) from a body of texts (corpus). It also allows users to search for those terms, returning the paragraphs where they are found in order of relevance.

Figure 1 shows the compound terms identified in a corpus of 112 legal documents in English, from the European Lynx project (<http://lynx-project.eu/>), on which the system was tested.

Figure 1. Example of compound terms identified in a corpus of legal documents, ordered by number of appearances (frequency), from the most frequent (at the top) to the least frequent (at the bottom).

Searching by compound term makes it possible to identify the locations in the corpus by means of **dispersion graphs** like the one shown in Figure 2. The more terms, whether compound or simple, that are added to the search, the fewer vertical lines there will be in the dispersion graph and, therefore, the fewer paragraphs there will be to search in to find the answer to the query.

Figure 2. Example of a *dispersion graph* for the term 'supreme court' in the corpus described.

The solution uses natural language processing technologies developed to process texts in different languages, as well as artificial intelligence and deep learning technologies. It has been developed using texts in Spanish, Catalan and English.

Business needs / application

- Searching for information has become a regular activity in our daily life. And not only in the work setting; in our leisure time we also look up where to see a film, where to go out to dinner, or which schools are near to where we live.

- We have all internalised the ‘Google style’ of searching for information: keywords. But there are situations in which we long for Google to understand the semantics of questions. For example, if we ask about ‘books that quote books by García Márquez’, Google will return links to websites containing books by García Márquez, instead of books that quote books by him. Google does not understand questions; it does not understand their semantics.

Competitive advantages

- Multilingualism: it works with texts in Spanish or English. Prototypes have been developed with other languages, such as Catalan.
- Intuitive searching for the terms most frequently used in a corpus of documents. For example, in a corpus of aeroplane repair manuals, there are terms such as ‘circuit’ or ‘gear’. A search for any of those terms would show the documents in which they are found. However, when the corpus is large, or when the term appears very frequently, searching by keywords becomes unfeasible.
- It proposes a search based on **compound terms**. Following the above example, compound terms would include ‘circuit breaker’ and ‘landing gear’. The user would type some letters of a term (not necessarily the initial letters) and the system would show the simple and compound terms present in the corpus, helping the user to create a more precise query. Searching by compound terms produces more accurate results than a traditional keyword (Google style) search.
- It offers linear scalability (using the product KeyQ-solr) and can be cloud-based (it has been placed on Azure).
- Protected by a software registration in the autonomous region of Madrid.

Can’t find what you’re looking for in a sea of documents? Our technology streamlines your searches by automatically identifying compound terms.

Past performance references

- The terminology generated by this tool is fed into KeyQ-solr, a system developed between 2020 and 2021 under the auspices of UPM’s Artificial Intelligence R&I Centre (AI.nnovation Space), a joint UPM-Accenture centre. Prototypes have been presented to the Spanish Ministry of Justice and the Government of Catalonia.
- The research group has expertise in semantic web and linked data technologies, well-established standards endorsed by international standards bodies such as the W3C. These technologies allow us to process questions semantically and provide more accurate results.

Protection

- **Software register:** M-007053 (March 2019)

Stage of development



Contact

KeyQ contact

Mariano Rico

ETSI Informáticos UPM, Ontology Engineering Group (OEG)

e: mariano.rico@upm.es