

IBERIAN ACADEMY DICTIONARIES AS LEXICAL RESOURCES

Ana Salgado, NOVA CLUNL 2019, Lisbon, Portugal

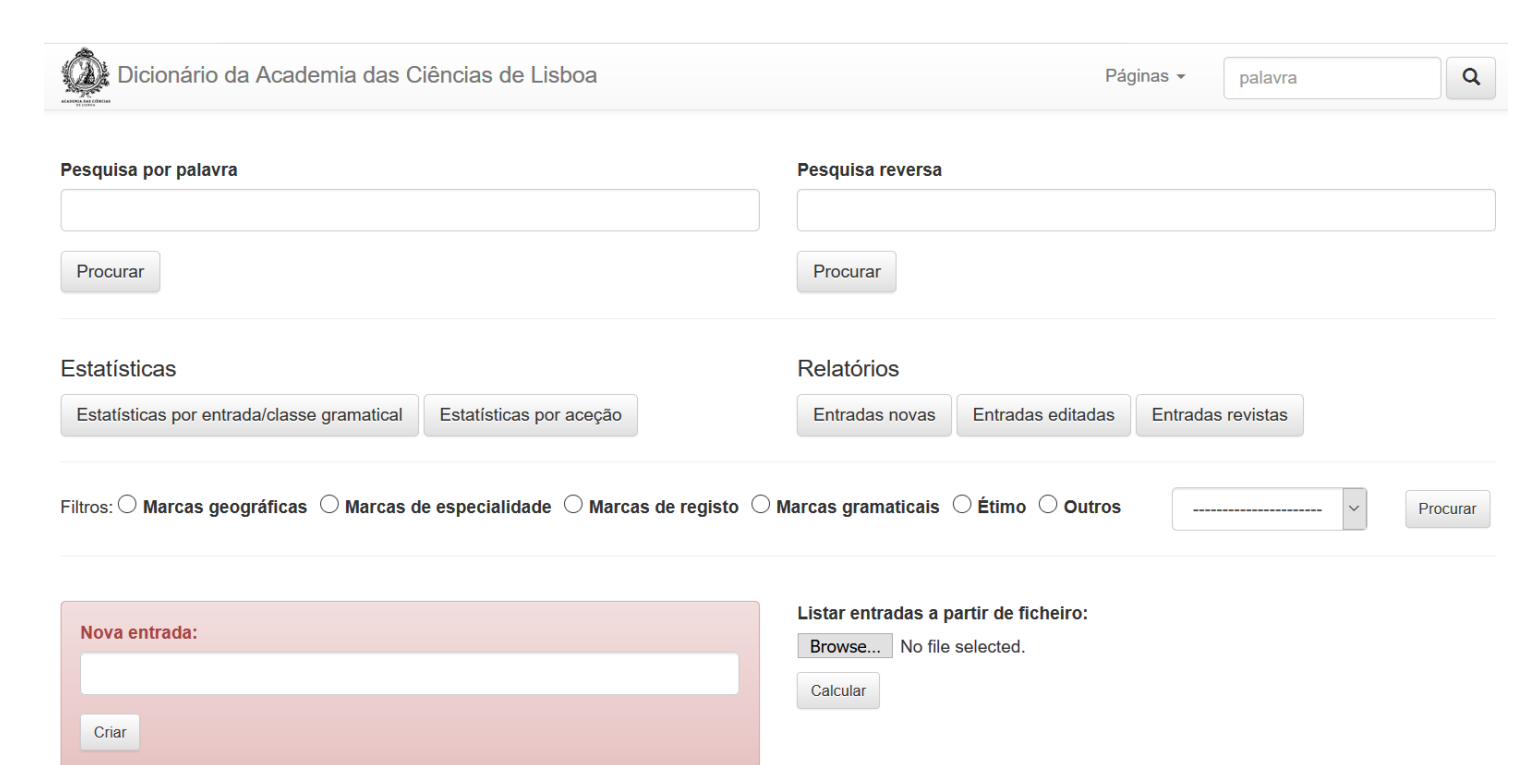
Introduction

The production processes of lexicographic work are changing to adapt to the digital era. To respond to the needs (users, interoperability purposes, data structure, consistency), standards have the advantage to facilitate the interoperability.

Portuguese Academy Dictionary (DLP)

The only complete print edition of a Portuguese Academy Dictionary was published in 2001. With the purpose of enlarging the dictionary, correcting imprecisions and mistakes, and making the dictionary available on the web, we created a database and developed a back-end system to manage the revision, creation and maintenance of the lexicographic service interface. Through a process of reverse engineering, the PDF was converted into XML files annotated using Text Encoding Initiative (TEI) guidelines. The resulting files were imported into an XML aware database (eXist) and are currently under review. To create, edit, delete and validate entries, we are using Oxygen XML Editor. Though not publicly available yet, the lexical data are annotated with elements (attributes) that are suitable for NLP purposes.

Simões, Alberto, José João Almeida, & Ana Salgado (2016). "Building a Dictionary using XML Technology". In *5th Symposium on Languages, Applications and Technologies*.



Goals

- To analyse the macro and microstructure of the *Diccionario de la lengua española* (Real Academia Española) and compare it with the DLP, especially to rethink the theoretical and methodological methods of the lexicographical tradition regarding usage labels that identify specialized lexicon
- To combine lexicographical and terminological methodologies
- To create a methodology for the selection, description and/or definition of linguistic and conceptual information (ISO TC37)
- To improve the macro and microstructures of lexicographical resources
- To increase the quality of lexical databases

Research issues

- How do lexicographers include terms in Iberian Academy Dictionaries?
- What is the percentage of terms in Iberian Academy Dictionaries?
- What do domain labels tell the user? Do they indicate a technical word? Are they useful?
- Which terms can you find in Iberian Academy Dictionaries? Are they all marked?
- How can you represent conceptual information?

Domain labelling in Iberian Academy Dictionaries

- How many domains are represented? What are the most common ones? Are they different? What are the most frequent ones?
- Is there a systematic label usage?
- Are there any recent domains that do not appear?
- What are the criteria for term selection and treatment?

Proposal for an Iberian agreement

- Proposal for an agreement between academies leading to a systematic labelling of the specialized use of a particular entry/meaning and its representation.
- In the digital age, a common and open line is needed to optimize the domain labelling process and its encoding in Iberian Academy Dictionaries, which are seen as structured, organized, standardized, accessible, and interoperable lexical databases.

Keywords: Lexicography, lexical database, dictionary, domain labelling, term, Terminology.



ELEXIS Visiting grants: Call 1
 Researcher: Ana Salgado
 Hosting institution: ELEXIS-ES: Real Academia Española (RAE, Spain), Centro de Estudios de la RAE y de la Asociación de Academias de la Lengua Española
 Institutional affiliation: NOVA CLUNL, Centro de Linguística da Universidade NOVA de Lisboa, Portugal
 Current position: PhD student in Terminology
 Period of stay: November 11-30, 2018

Connect with us!



<https://www.facebook.com/clunl>



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015.