



Medical Semantic indexing in Spanish

<http://temu.bsc.es/mesinesp>



Plan TL

Plan de Impulso de las
Tecnologías del Lenguaje



The challenge Medical Semantic indexing in Spanish, Task A Spanish

Task A Spanish will be based on the standard process followed by **LILACS** and **IBECS** to index journal abstracts in Spanish.

▶ The participants will be asked to classify new documents, written in Spanish, as they become available on-line, before **LILACS** and **IBECS** curators annotate them manually.

▶ As manual annotations become available, they will be used to evaluate the classification performance of participating systems.

TASK AWARD, the Plan TL is sponsoring the challenge: 1,000€ to winners and 500€ to the second classified teams.

Motivation

Efficient access to medical literature is a pressing need not only for information published in English but also for articles published in other languages. Efficient retrieval of medical publications is key for evidence-based medicine, preparing systematic reviews or finding particular clinical case studies. Query expansion approaches relying on indexing with structured vocabularies is an efficient approach to facilitate more powerful literature search engines.

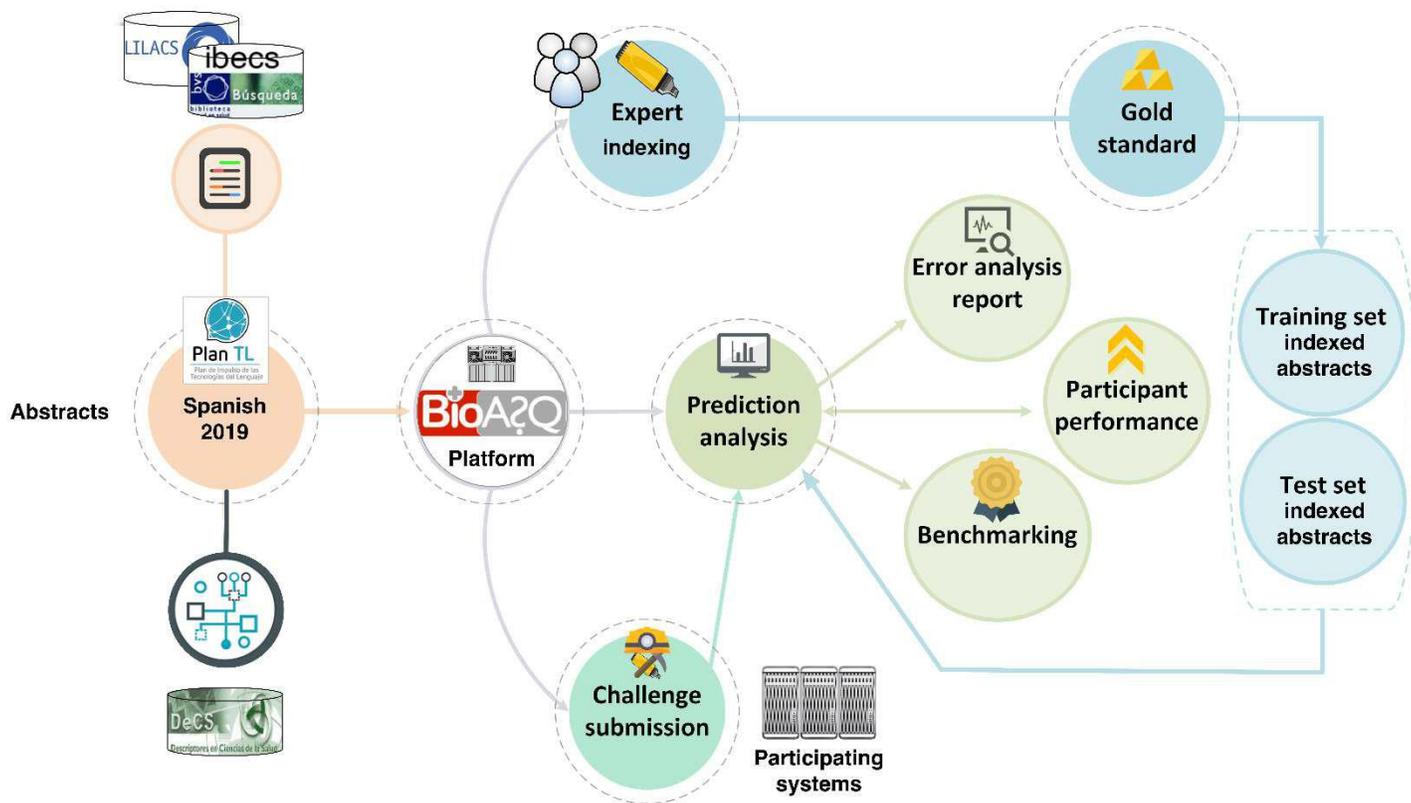
The critical importance of semantic indexing with medical vocabularies motivated several-shared tasks in the past, in particular the **BioASQ** tracks, with a considerable number of participants and impact in the field.

Currently, most of the Biomedical NLP and IR research is being done on English documents. Nonetheless, it is important to note that there is also a considerable amount of medically relevant content published in other languages than English and particularly clinical texts are entirely written in the native language of each country, with a few exceptions.

There is a large subset of medical content published in Spanish each year. Resources like PubMed do only contain a fraction of the biomedical and medical literature originally published in Spanish, which is also stored in other resources such as **IBECS**, or **LILACS**.

Following the outline of previous medical indexing efforts, in particular the success of the BioASQ tracks centered on PubMed, we propose to carry of **the first task on semantic indexing of Spanish medical texts**.

Thus this task will address the automatic indexing with structured medical vocabularies (DeCS terms) of abstracts from the IBECS and LILACS databases written in Spanish. The main aim is to promote the development of semantic indexing tools of practical relevance of non-English content, determining the current-state-of-the art, identifying challenges and comparing the strategies and results to those published for English data.



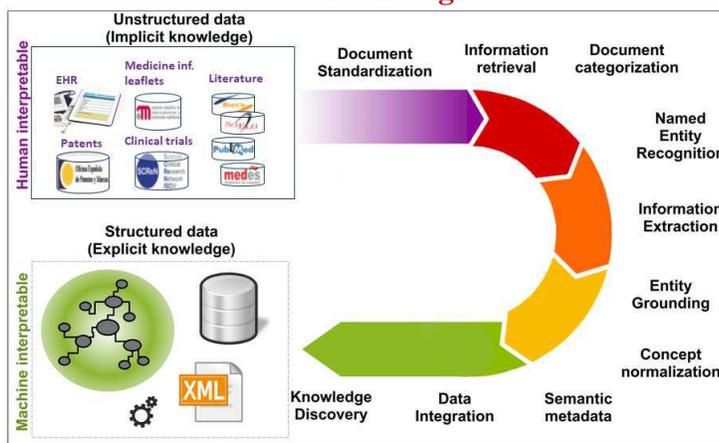
Shared tasks organization

Plan TL

The **Plan TL** aims to promote the development of NLP and machine translation in Spanish and Spain's co-official languages. One of the flagship projects of the Plan TL is related to Healthcare and Biomedical domain. The objectives of the plan are

- Increase the amount, quality and availability of **linguistic infrastructure**.
- Transfer knowledge from the research field to the industry.
- Improving the quality and capacity of public services, integrating NLP and machine translation technologies. Identify **use cases in public administration**

Text Mining



Shared tasks: the driving force behind LT (language technologies)

Shared tasks are a way to promote research while being an important step towards standardization in key aspects such as shared formats and evaluation metrics and criteria. They involve the development of new resources.

