

An update from Dialnet in the Spanish Research Ecosystem

Setting the scene

Non-profit Public Foundation created and attached to the University of La Rioja

Born in 2009 to manage and develop Dialnet Repository (the largest database of Spanish scientific production in the World).

Collaborative and integrative nature (efforts – resources – services)

Dialnet is core of a strong collaborative network of institutions in Spain (65 Universities collaborating among others). Most of the Spanish Universities are taking part in the project and the system is becoming a reference in the Spanish research community.

As the largest repository and infrastructure provider of Spanish scientific production, Dialnet is also a key player in building bridges between Europe and Latin American countries (57 Universities already collaborating).

What does it bring to the table?

Over time, Dialnet ambition is trying to achieve hollistic solutions for the whole research scenario. For that purpose, Dialnet id providing a wide range of services to the Spanish research community. The main services are:

Dialnet Plus

Dialnet CRIS (at an institutional, regional and multiinstitutional level)

Dialnet Metrics

Dialnet Data

...

These resources follow the Open Science principles and are designed to be interoperable with other systems for a global research ecosystem.

AI ongoing projects

- Managing and recognizing authorship through knowledge objects formation (GRAIN): Research outputs interconnection, relevant documents or authors, fraud detection in references.
- Classification and revolutionary analysis through embeddings and semantics (CARES): Multi-vectorial database, Zero-shot LLM for NER tasks, Zero-shot capabilities for classification, Integration of the system in Dialnet CRIS and Web interface and API.
- Visualization of the life cycle of an idea/research topic: evolution, influence and generation of new lines (VICVID): Classification and sentiment of citations, semantic analysis of publications, visualization

AI showcase: A Multidimensional Knowledge Map of the Spanish Research Scenario

The scope of the paper was to show a proof of concept of a multidimensional knowledge map (not a force-directed graph) built with state of the art AI techniques.

To build the map we used a pre-trained model for sentence embeddings (Sentence Transformers) and a pre-trained model for topic extraction. The map is built using UMAP, a dimensionality reduction technique to be able to show using the 3D coordinates of the documents.

We believe this is a novel technique to build a knowledge map, as it is not based on predefined relationships among the documents, but on the semantic similarity/proximity (cosine, euclidean, etc...) of the documents obtained using the same technology that powers modern Large Language Models.

This solution is a step forward in the field of knowledge maps, as it allows to uncover hidden knowledge without predefined connections among the documents.

We include some references to the innovative technologies used in the paper:

- Obtaining an embedding using title+abstract and Sentence Transformers (model: <https://huggingface.co/sentence-transformers/paraphrase-multilingual-mpnet-base-v2>).
- Reference to the paper on transformers: "Attention Is All You Need" (<https://doi.org/10.48550/arXiv.1706.03762>).
- Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks <https://doi.org/10.48550/arXiv.1908.10084>
- Dimensionality reduction using UMAP to 3 dimensions (<https://doi.org/10.48550/arXiv.1802.03426>).

Latest developments

CERIF-XML export of Dialnet CRIS data (PCR integration)

Automatic generation of reports for sexenios and academic reports

Funding and projects released in Dialnet CRIS for a transparent and open research ecosystem

Addition of FECYT badge and APC information in Dialnet CRIS

Other achievements

Working Groups launched under the umbrella of the Spanish CRIS Network (REBIUN)

Official presentation of the Regional Collective CRIS of Castilla y León.

Dialnet Metrics has been officially included in the ANECA accreditation process for Spanish Universities.

SDGs showcasing in Dialnet CRIS

Future steps

CVN (National Researcher Curriculum) platform

Research Data Repository connected to Dialnet CRIS

OpenAlex integration in Dialnet CRIS

Infrastructure and Instruments module in Dialnet CRIS

Expanding the collaboration network in Latin American countries

Participation in the AI4CRIS Working Group of euroCRIS to share our experiences