.

Medieval Iberia through Pelagios Commons.

María Gimena del Rio Riande.

Cita:
María Gimena del Rio Riande (2016). *Medieval Iberia through Pelagios Commons.*

Mediaeval Iberia through Pelagios Commons

The LINHD team at UNED (http://linhd.uned.es/) has been awarded one of the precious Pelagios grants for a proposal to extend Pelagios to the Mediaeval Iberian world, identifying placenames from a very important source for Spanish Philology: the Biblioteca Digital de Textos del Español Antiguo (BIDTEA, Hispanic Seminary of Mediaeval Studies-University of Wisconsin-Madison) (http://www.hispanicseminary.org/manual-es.htm)

In this stage of the project we will concentrate on the Documentos Castellanos de Alfonso X. The texts are edited and revised by us following the *Textos y Concordancias Electrónicos de Documentos Castellanos de Alfonso X* by Mª Teresa Herrera, Mª Nieves Sánchez, Mª Estela González de Fauve and Mª Purificación Zabía (Madison: Hispanic Seminary of Medieval Studies, 1999). The regions to which we will devote our work in the gazetter are: Andalucía, Galicia, León, Murcia, Castilla la Nueva and Castilla la Vieja.

The project focuses on two main objectives:

1. Contribute to extend Pelagios core by adding Mediaeval Iberian placenames.

2. Develop a framework for knowledge extraction of Mediaeval Iberian placenames in reliable and verifiable digital sources in Spanish.

To achieve our goals, some important issues have to be addressed. Mediaeval Iberian placenames are not easy to identify and locate for several reasons. Firstly, there is not a Spanish/Iberian gazetteer to identify geographical places from literary sources and its equivalents in the current geography. Moreover, it is also necessary to regularize names and terms, as Mediaeval surviving texts contain many phonetic and linguistic variations to make name-entity recognition technologies (NER) systems recognize them automatically. Finally, some locations that appear in literary texts were invented by Mediaeval authors, they never existed or were named in a different way throughout the centuries.

For this purpose, a Mediaeval Iberian knowledge extraction framework will be developed to implement the different phases defined in a knowledge extraction procedure:

1) Adquisition and identification

We will explore datasources of old Spanish geographical names. To acquire the set of Mediaeval placenames a harvesting procedure over the digital sources will be implemented. The reliable and verifiable digital sources are the mentioned ones of the Biblioteca Digital de Textos del Español Antiguo (BIDTEA, Hispanic Seminary of Mediaeval Studies -University of Wisconsin-Madison).
We will also test different Natural Languages Process (NLP) tools to extract automatically geographical entities with NER systems. Digital NER tools like Freeling (http://nlp.lsi.upc.edu/freeling/) will be applied especially adapted to the Old Spanish variants.

2) Knowledge creation

Once the Mediaeval Iberian placenames are recognized and disambiguated, they will be added to Pleiades Gazetter. All the new information about Mediaeval Iberian placenames should be categorized and indexed semantically as URIs that could be reused and linked to other data models into a linked open data context (data models from National Spanish Library, digitized corpora, literary works, dictionaries, wiki sources…).

3) Evaluation

We will use Pelagios as a research framework to test the system created by entering texts in the Recogito API in order to check if the new functionalities of the installed vocabulary get the expected results and the Mediaeval Spanish places mentioned in the texts are correctly recovered. For this step, we will use different type of texts and ask different users from our community to try the API with different texts.

Finally we will mark the texts in Recogito.

We hope our project may also serve as a pilot for a larger initiative in relation to the inclusion of Mediaeval Iberian data in Pelagios Project. We think researchers interested in Mediaeval Iberian texts and GIS will stronlgy benefit from this resource and become more engaged with the use Pelagios Commons and its Community.
We have set a starting point for the project, first in our DH Summer School at UNED, Madrid, were Pau de Soto gave us a workshop about Pelagios, and at the I National Digital Humanities Conference in Argentina (7-9 November 2016) were Gimena del Rio and Melisa Martí will also teach how to work in Recogito using the texts of this project as example (http://www.aacademica.org/aahd.congreso/tabs/register).

Our team is interdisciplinary and digital humanities focused, as it is composed by a humanities team with experts in Mediaeval Spanish Literature, and a Computer Science one with high expertise in data analytics and data processing working at LINHD, the Digital Humanities Innovation Lab at UNED, a DH center focused on semantic web technologies and Spanish literature. This group is working together in several projects related, such as POSTDATA (Poetry standardization and linked open data), or EVILINHD (a virtual research environment for digital humanities research), which are strongly related with the proposed resource and Pelagios Project.

We are right now working hard in the project and we will be grateful for any comments and suggestions!

Gimena del Rio Riande (CONICET, LINHD UNED)

Elena González-Blanco García (LINHD UNED)
Clara Isabel Martínez Cantón (LINHD UNED)

Ma. Jesús Redondo Rodríguez and all the LINHD team