

Alesson Ramon  
Rota

## Data Mining as Analytics for Research in History

This communication proposal will show some of the methods used for research in digitized collections produced from physical documents (paper). That is, how to do a big data search from historical documents. As an example digital collections for online consultation will be used, such as the Hemeroteca Nacional (Brazil), the Ibero-Amerikanisches Institut (Germany) and the Archivo General de Indias (Spain). The methods originate from the relationship between computing and the humanities, sometimes discussed as digital humanities. Techniques for downloading large amounts of information (called web scraping), will be discussed, as well as algorithms to convert the images of physical documents into digital text. Data meaning techniques are used to find what is considered relevant, making it possible to build new forms of data representation and analysis.

Keywords: Digital Humanities, Digitization, Data Mining, Image Processing