Most of the current musicological knowledge is present in printed books and manuscripts. In the last years greats efforts have been done in order to digitize and make available these documents in form of Digital Libraries. However, digital documents are mainly stored as raw text, with no more structure than indexes and some metadata. Therefore, implicit knowledge contained in text is not understandable by computers and cannot be processed like that.

Automatic processing of text documents may help musicologists in several ways, such as improving navigation through a library, discovering hidden knowledge, accelerating tedious tasks, etc. To apply these techniques to a Digital Library, the information contained in documents should be carefully structured and semantically annotated. Information Extraction is a discipline of computer science focused on the extraction of structured information from unstructured text sources.

We propose a method to automatically extract meaningful knowledge from documents present in Digital Musical Document Libraries, by using Information Extraction techniques. Our method has two main steps. First, relevant named entities (e.g. composers, organizations, places, etc.) are identified in the text. Second, words between these entities are syntactically and semantically analyzed to understand the relationship between them. Finally, the extracted knowledge is represented in a machine-readable format as a knowledge graph, where entities are represented as nodes, and relations as edges.

The resulting knowledge representation is finally visualized as an interactive graph. With the proposed information visualization, users may go from one document to another by browsing the knowledge graph. We tested our method with a subset of artist biographies present in the Grove Music Online.