

Citation mining of humanities journals: a comparison of the projects Cited Loci and Linked Books

Giovanni Colavizza and Matteo Romanello

Digital Humanities Laboratory, École Polytechnique Fédérale de Lausanne

Abstract

Even large citation indexes such as the Web of Science, Scopus or Google Scholar cover only a small fraction of the humanities. This coverage sensibly decreases going backwards in time. Citation mining of humanities publications – defined as an instance of bibliometric data mining and as a means to the end of building comprehensive citation indexes – remains an open problem. In this contribution we discuss the results of two recent projects that have dealt with mining citations from humanities journals: Cited Loci and Linked Books. The former focused on the domain of classics, using as a corpus the articles in JSTOR; the latter considered the historiography on Venice and a novel corpus of journals and monographs. Both projects attempted to mine citations of all kinds, e.g. abbreviated and not, to all types of sources, including primary sources, and considering a wide time-span (19th to 21st century). We first discuss the current state of research in citation mining of humanities publications. We then present the various steps involved into this process, from corpus selection to data publication, discussing the peculiarities of the humanities in these respects. The approaches taken by the two projects are compared, allowing us to highlight disciplinary differences and commonalities, as well as shared challenges between historiography and classics on this respect. In closing, future perspectives and open problems are sketched.