Versão integral disponível em digitalis.uc.pt

Materialidades da Literatura

Vol. 4.2 (2016)
ISSN 2182-8830
‘Estudos Literários Digitais 2’
Manuel Portela e
António Rito Silva (orgs.)
Annotation of Toponyms in TEI Digital Literary Editions and Linking to the Web of Data

FRANCESCA FRONTINI
Istituto di Linguistica Computazionale “A. Zampolli”, CNR, Pisa

CARMEN BRANDO
EHESS, CRH UMR 8558 (EHESS-CNRS)

MARINE RIGUET
Labex OBVIL – Université Paris-Sorbonne

CLÉMENCE JACQUOT
Université d’Artois, EA Grammatica

VINCENT JOLIVET
Labex OBVIL – Université Paris-Sorbonne

Abstract
This paper aims to discuss the challenges and benefits of the annotation of place names in literary texts and literary criticism. We shall first highlight the problems of encoding spatial information in digital editions using the TEI format by means of two manual annotation experiments and the discussion of various cases. This will lead to the question of how to use existing semantic web resources to complement and enrich toponym mark-up, in particular to provide mentions with precise georeferencing. Finally the automatic annotation of a large corpus will show the potential of visualizing places from texts, by illustrating an analysis of the evolution of literary life from the spatial and geographical point of view. Keywords: digital literary studies; toponyms; semantic web; geographic databases; maps and visualizations.

Resumo
Este artigo aborda as dificuldades e as vantagens da anotação dos nomes de lugar em textos literários e de crítica literária. Começamos por realçar os problemas de codificação da informação espacial em edições digitais usando o formato TEI, através de duas experiências de anotação manual e da análise de diversos casos. Isto conduzirá à questão de como utilizar os recursos da web semântica para complementar e enriquecer a marcação de topónimos, em particular com georreferenciamento rigorosa. Por último, a anotação automática de um grande corpus irá mostrar o potencial de visualização de locais a partir de textos, ilustrando a análise da evolução da vida literária segundo um ponto de vista espacial e geográfico. Palavras-chave: estudos literários digitais; topónimos; web semântica; bases de dados geográficas; mapas e visualizações.
1. Introduction

We propose here an excursus into the interactions between literary studies and geographical information science. In particular, we shall examine the issue of correctly and efficiently annotating place names in literary texts and (literary) criticism. Clearly, the problems connected to these two tasks are similar but not identical. Literary texts may contain fictional places, while non-fictional texts mostly real ones. Also the goals of such annotations are different; in the first case the purpose of the annotation may be to study the interaction between setting and narrative space\(^1\) within a single text, while in the second case it will often be to enable a diachronic analysis of large corpora in order to find trends and evolutions in the geographical distribution of literary centres and topics. Nevertheless, many similar problems arise, and it is useful to address the common issues of toponym annotation in the literary domain in a comprehensive way, as we shall do in this paper. We shall argue that some reflection is necessary to establish best practices for the appropriate annotation of place names in texts and for their linking to existing geographical databases to be able to retrieve information—typically but not exclusively geospatial information—about them.

The keen interest of researchers in the Humanities for the geographical dimension of information is not a recent phenomenon. The possibility of modelling, storing, analysing (via spatial analysis methods) and visualising geospatial information proposed by Geographical Information Systems (GIS) have been exploited particularly by archaeologists and by historians to study, trace and quantify phenomena taking place on the surface of the Earth. More recently, disciplines within the Digital Humanities (DH) have shown increasing interest in geospatial information. As a proof of this, a Geohumanities special interest group\(^2\) is particularly active within the DH community. The availability of geographical databases such as interoperable gazetteers thanks to Linked Open Data (LOD) initiatives as well as the interactive cartography tools and technologies offered by the Web represents an undeniable opportunity for these communities to explore novel interdisciplinary research ideas.

For what concerns (digital) literary studies, some interesting projects have recently seen the light, inspired in part by the pioneering work of Franco Moretti (Moretti 2007) and Matthew Jockers (Jockers 2013). In general, these are large projects requiring interdisciplinary work among literature researchers, corpus linguists, geographers and cartographers, and others; at the same time, they allow researchers to gain new insights from working on large sets of data and from aggregating them in novel ways, which allows new

---

1 For the theoretical complexity of analysing fictional spaces and for a proposal of visualisation of several spatial dimensions in fiction see Hones (2011) and Piatti et al. (2013).