

**Abstract for the Spatial Humanities
International Conference**

Article Title: *Experiments in joining GIS and NLP for literature in Portuguese*

Authors: *Daniel Alves, Diana Santos, Danielle Sanches and Ângela Matos Pité*

We intend to describe the BILLIG initiative, a project funded by EEA Grants programme, for fostering cooperation between Geographic Information Systems and corpus linguistics approaches to literature. The project makes use of *Atlas das Paisagens Literárias*, an SQL database that at the present comprises around 7500 excerpts from almost 400 works which were human-annotated by 49 people on many aspects of landscape and environment in Portuguese literature, and *Literateca*, a corpus environment with more than 800 full length works from Brazil and Portugal (mainly novels but also non-fiction works) by more than 200 authors with places classified but not geolocated (and possibly including many fictional places). Through the sharing of knowledge between both teams and the attempt to analyze in a comparative way both methodologies for annotating literary texts, we will report on two pilot tasks for works of the 19th and 20th centuries: first, a comparison of the *Atlas* human annotation, as far as locations are concerned, with the output of the automatic named entity system used by *Literateca*, to establish advantages and limitations of both procedures; second, add geolocations to *Literateca*'s place names and develop a Web interface to display the results, using *Atlas* previous experience. We will also discuss the distribution of place names in Brazilian and Portuguese literature (some authors are regional, other global), the frequency of reference to places and the use of imaginary versus real places.