



## **Text Data Mining in a Geospatial Metadata Catalog**

Fátima Ferrão dos Santos\*

Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, Brazil –

[fatima.santos@ibge.gov.br](mailto:fatima.santos@ibge.gov.br)

### **Abstract**

The characteristics that make geospatial data special as a computing problem have been acknowledged in many ways and one of the mainly important initiatives is the development of Spatial Data Infrastructure (SDI). This term refers to the integrated set of technologies, standards, policies, institutional arrangements and human resources to facilitate the availability, access and use of geospatial data and information to support policy, business, research, govern and society at large. In order to share the geospatial information, it is necessary to deploy metadata. Mining metadata can enable the extraction of patterns and useful and new knowledge in the context of this application. The aim of this paper is to present a methodology to implement text data mining in a geospatial metadata catalog. The focus of this paper is not on developing the spatial data infrastructure but on developing data mining tasks within the emerging infrastructure. To assess this information, a network approach based in probability estimative was used. These techniques are aimed at discovering natural divisions of networks into groups, based on metrics of strength of connection between vertices. Comparisons among the generated networks are made in the light of different categories of metadata, the closeness and betweenness centralities measurements. Our results show the main pattern among distinct categories and how keywords are correlated to these main categories. We conclude that the approach of complex networks is a promising tool for metadata analysis.

**Keywords:** text data mining; metadata; spatial data infrastructure; networks.