Disentangling a non-standard language corpus by steering visualizations to answer questions

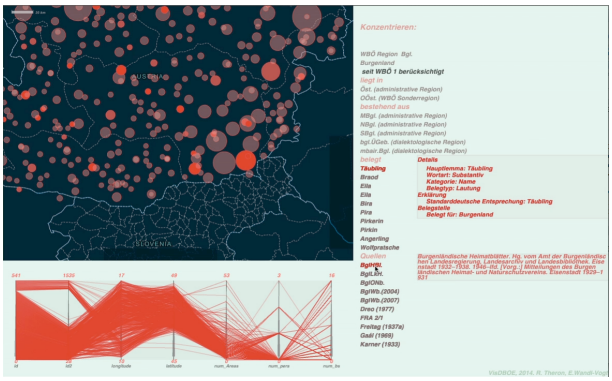
EVELINE WANDL-VOGT
Austrian center for digital humanities & Austrian academy of sciences

ROBERTO THERON
Department of computer science and automation - University of Salamanca

Similarly to what has happened in all domains dealing with vast amounts of information, dictionaries can greatly benefit from a visual access in order to grasp the inherent tangled and complex nature of the knowledge encapsulated in them.

In the past, there have been several visualization efforts mainly aimed at escaping the idea of dictionaries as flat lists of entries. The use of hyperlinks and graph drawing techniques [1] are the main contributions of the past decade. While this is an advantage, several aspects remain that can dramatically change the users' experience: an enhanced access to the wealth of sources related to dictionaries tailored to the particular needs and tasks of such users can indeed surpass this very notion of a flat dictionary.

This previous statement is especially relevant when non-standard data are considered. Space and time become the steering wheel of any inquiry. In this regard, there are few works that take full advantage of visual interactive access to non-standard dictionaries. In one of the first efforts in this direction can be found.

Language has been addressed from a Geographic Information Systems approach very early, incorporating later advanced computational linguistics features and fields such as dialectology and dialectal lexicography. However, dictionary publishers have not followed this approach very often.

Thus, the main motivation of our work is to provide a non-conventional access to a digital non-standard dictionary, where the user is encouraged to explore the data and discover patterns or gain unexpected knowledge by interacting directly with different views, and embed it as European research infrastructure.