Eurographics Conference on Visualization (EuroVis), Posters Track (2017)
A. Puig Puig and T. Isenberg (Editors)

Structuring and visualizing knowledge in soil science

B. Stößel[1], E. Rabot[1], F. Richter[2], U. Weller[1] and H.-J. Vogel[1]

[1]Helmholtz Centre for Environmental Research - UFZ, Department of Soil Physics, Halle (Saale), Germany
[2]Helmholtz Centre for Environmental Research - UFZ, IT Department, Leipzig, Germany

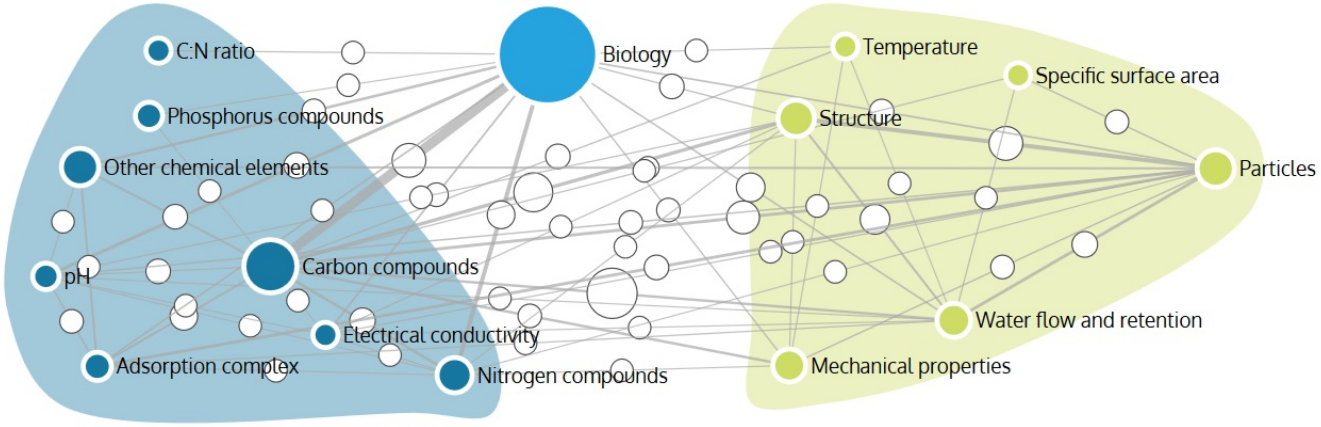

Figure 1: *Soil property-centered relationship map. It shows a dynamic graphical representation of the mySQL-database contents. The colored bubbles represent the soil properties listed in a hierarchical keyword tree. White bubbles represent groups of scientific articles dealing with the properties connected to them. The bigger the bubble, the more articles are connecting both properties.*

Abstract

Soil is a highly complex system. Myriads of articles are published each year on soil processes and properties, interactions between them and management practices influencing them. But, because of the sheer mass of information and the complexity of the system, it is hard to maintain an overview of reliable information on these interactions and their sensitivity to external forcings. To collect, structure and visualize scientific results on functional relationships, we develop a system based on a mySQL database, a soil thesaurus and two JavaScript-driven visualizations. The system is fueled by a comprehensive soil literature review and currently being organized as a community effort to improve the understanding of the soil system and support users in finding information on functional relationships and knowledge gaps in soil science.

Categories and Subject Descriptors (according to ACM CCS): J.2 [Computer Applications]: Physical sciences and engineering—Earth and atmospheric sciences E.1 [Data]: Data structures—Graphs and networks H.3.4 [Information systems]: Information storage and retrieval—Systems and software - Information networks

1. Introduction

The soil system is highly complex. This is underlined by the amount of publications on soil-related topics published each year. These articles deal with soil processes and properties (variables), interactions between them and how they are influenced by external forcings (e.g. management practices) under specific side conditions. But, because of the sheer mass of information and the complexity of the system, it is hard to maintain an overview of reliable information on these interactions and their sensitivity to forcings. Hence, it would be highly valuable to have a library for structured soil knowledge to be able to support and improve system understanding. Common tools to search for soil related knowledge include search engines like "Google Scholar" or "ISI Web of Knowledge". But, these tools only allow searching for articles by scanning for specific words or text passages in title, topic, article texts or user

DOI: 10.2312/eurp.20171154

defined keywords. With the tools described, it is not possible to get an easy overview of functional relationships between soil properties or processes under specific side conditions. To close this gap, we designed a mySQL database, fed with data on soil processes and connections between important biological, physical, and chemical soil properties gained by reviewing a large amount of scientific literature. The contents are presented by two JavaScript-driven visualizations. They can be used to search for specific publications or to explore the soil system by means of an interactive relationship network to identify knowledge gaps.

2. Database & Visualization technique

A mySQL database combined with the Django framework has been designed to collect data on soil research results extracted from the literature (Figure 2) or entered directly by the authors of scientific papers. The database was normalized to reduce data redundancy and improve integrity. Researchers can enter the major findings of their study in a web form, submitting it directly to the database. To create a new data entry, users have to enter bibliographic information first. After selecting the side conditions relevant for the study, soil variables and interactions between them can be specified. Interactions and side interactions are functional relationships between soil variables, recorded as types of functions (e.g. linear, unimodal, saturation). Users can choose from a set of function types, common in soil science, to efficiently describe their major findings. During the process of data entry, a newly created soil thesaurus is utilized to guide the user. The thesaurus is a hierarchical keyword tree based on soil properties, processes and management practices extracted from the reviewed literature. It allows the database to work with standardized wording for all relevant pieces of information entered. Users can expand it further by adding new keywords through a keyword creation module. The thesaurus will be updated including new keywords after a review process. The visualizations utilize the thesaurus to allow for an easy creation of functions to filter database contents (e.g. showing only interactions in sandy soils). They are using the graphical JavaScript library D3 (Data-Driven Documents) [BOH11] to create interactive relationship maps. The property-centered visualization (Figure 1) is a map of important soil properties and processes connected by scientific publications [Bon17b]. The user can start the search at a more general level of keywords and can then zoom into greater details following the hierarchical thesaurus. This can be done directly in the relationship map or by using the graphical representation of the hierarchical keyword tree. A list of articles, including bibliographic information and doi-links, dealing with the relationship of interest can be accessed by clicking the white bubbles between the properties. All properties are draggable and a zoom function is integrated, allowing the user to rearrange the visualization according to his needs. The article-centered visualization (Figure 3) is a network of articles connected by soil properties, processes and management practices. It uses a central article as starting point and invites the user to discover articles related to the central article by a common soil property, process or management practice [Bon17a]. To maintain a good overview, only articles within a distance of two nodes from the central article are shown. Articles close to each other in terms of shared soil properties are dragged towards each other by a gravitation force to demonstrate their closeness.

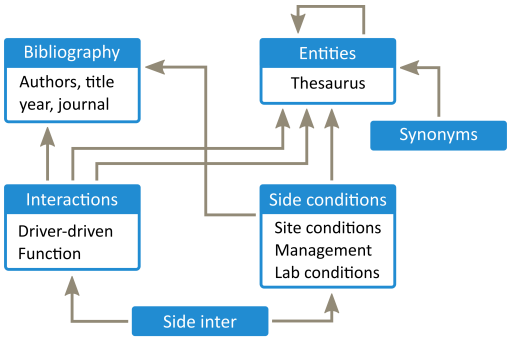

Figure 2: *SQL database structure.*

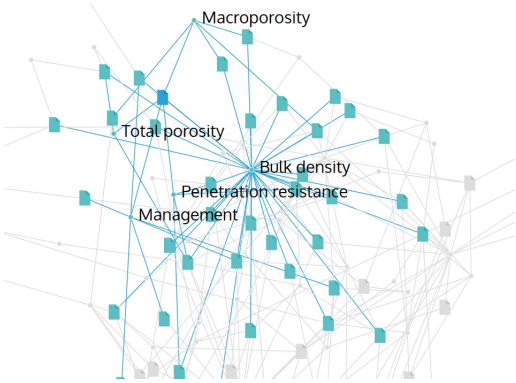

Figure 3: *Article-centered relationship map. The page-shaped symbols represent articles in the database. The symbol color represent their distance from the central article (blue = central article (CA), turquoise = sharing at least one property with the CA, grey = maximum 2 nodes away from the CA). Dots represent soil properties, processes and management practices, analyzed by the article.*

3. Future plans

The system we present is still work in progress and will improve further. The visualizations will provide information on the type of functional relationships. Database and thesaurus will expand by a community effort fueling the visualizations with more information. In additon, the use of a WebGIS-Tool will enable filtering research articles and functional relationships by spatial information (e.g. coordinates, climatic zones, etc.). Furthermore, it is planned to test automatization of the extraction of functional relationships from soil articles by text-recognition algorithms.

Acknowledgements

This work was funded by the German Federal Ministry of Education and Research (BMBF) in the framework of the funding measure "Soil as a Sustainable Resource for the Bioeconomy - BonaRes", project "BonaRes (Module B): BonaRes Centre for Soil Research, subproject A" (grant 031A608A).

References

[BOH11] BOSTOCK M., OGIEVETSKY V., HEER J.: D^3 data-driven documents. *IEEE transactions on visualization and computer graphics 17*, 12 (Dec. 2011), 2301–2309. doi:10.1109/TVCG.2011.185. 2

[Bon17a] BONARES: Article-centered bonares relationship map, Apr. 2017. http://bonares-litlib.ufz.de/visualization/paperMap.html?paperID=405. 2

[Bon17b] BONARES: Property-centered bonares relationship map, Apr. 2017. http://bonares-litlib.ufz.de/visualization/propertyMap.html. 2