Data and Text Processing for Sustainability

Francisco M. Couto[1][0000-0003-0627-1496]

[1] LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal
fjcouto@edu.ulisboa.pt

Abstract.

Environment Ontology and Environmental conditions, treatments and exposures ontology are two examples of the current interest of the scientific community in providing comprehensive and accurate knowledge bases that help us more effectively to navigate and retrieve information from literature, find evidence, integrate data, generate hypothesis, or even discern relevant from irrelevant data. However, most of these benefits requires an in-depth understanding of complex and sophisticated technologies, that are constantly evolving. This manuscript proposes an alternative solution that is available to every sustainability specialist without becoming dependent on continuous financial support, third-party applications, or advanced computer skills. This alternative may not offer the full state-of-the-art potential, but offers a feasible and efficient starting solution to explore the main potential of automatically processing data and text using knowledge bases. The solution is to understand the basics of shell scripting in order to perform named-entity recognition and linking by following the book, entitled Data and text processing for health and life sciences, which shows how we can process data and text the same way we conduct a laboratory protocol i.e. testing and understanding its multiple procedural steps, variables, and intermediate results. Additionally, a fully functional shell script, MER - Minimal Entity Recognizer, (available also as a web tool) can be explored to recognize concepts related to sustainability. The open source software is available at: https://github.com/lasigeBioTM; the web tools and all the open access book material at http://labs.rd.ciencias.ulisboa.pt/. As an example, we copied the titles of the keynote talks of this conference and selected the Environmental conditions, treatments and exposures ontology as lexicon. The results show not only the concepts found, such as hydrogen and biomass, but also their direct links to the respective entry in the ontology.

Keywords: Text Processing, Semantics, Ontologies.