From relational databases to Linked Data: R for the semantic web

Jose Quesada[1*]

Max Planck Institute, Adaptive Behavior and Cognition (ABC), Berlin Germany,
Quesada@gmail.com

Keywords: Semantic web, RDF, store, RDBMS, modeling

Most of the current structured data in the world lives in relational Data bases (RDB). But statisticians and practitioners sooner or later will find themselves dealing with a different kind of structured representations in the Resource Description Framework (RDF). The logical evolution of the current Web of documents into a Web of Data (and ultimately a Semantic Web) requires mapping of vast quantities of data from RDB to RDF. The conceptual foundations of the relational model and RDF are indeed quite similar, based as they are on set theory and relationships. However, there are important differences between RDBs –or XML- and RDF: in the former the schema often describes a tree, while the latter uses a (more general) graph. I will cover existing options to transform available RDBs into RDF. While there is currently no RDF library, R can interface with different existing frameworks, both for storage (on memory and disk) or processing (reasoning and querying). Then, I will present ways in which R can handle RDF data natively. I will use *package ff* for storage with its C++ core implementing fast memory mapped access to flat files, and *package igraph*, designed to deal with large graphs.

References
W3C Resource Description Framework (RDF)
    http://www.w3.org/RDF/
Adler, D., Gläser, C., Nenadic, O., Oehlschlägel, J., & Zucchini, W. (2008). ff: memory-efficient storage of large atomic vectors and arrays on disk and fast access functions.
Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. International Journal, Complex Systems, 1695.
Team, R. D. C. (2008). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from http://www.R-project.org.