

R4X : Simple XML Manipulation for R

Romain François – *Mango Solutions*

useR! 2008. Dortmund.

Abstract

Data transfer is an important component in many multi-technology applications. The eXtensible Markup Language (XML) is a medium of choice for exchanging various sources of data. Recent developpements at Mango Solutions have justified the production of an R package to provide convenient manipulation of XML structures.

Based on the powerful parsing facilities of the `XML` package[4] and templating abilities of the `brew`[3] package, `R4X` gives R users a simple mechanism to create, read and manipulate XML structures. The functionality of the package is conceptually based on the E4X[2] standard which promotes XML as a core data-type of the javascript language. In order to create a seamless integration of XML into R, much of the functionality of E4X has been ported to `R4X`.

`R4X` provides a convenient environment for the *creation* of XML structures, through the single generic `xml` function. `R4X` also features simple *manipulation* of XML structures via the usual R slicing operators (`[` and `[]`) combined with a syntax close to XPATH in order to extract arbitrarily nested content from an XML structure.

This presentation will describe key features of the `R4X` package and discuss anticipated extensions of the functionality. Examples will be used to demonstrate the use of `R4X` to build a simple Rich Site Summary (RSS) reader, generate a tag cloud of the description of current CRAN packages in xHTML, create Scalable Vector Graphics (SVG) and a custom RUnit[1] protocol report based on the Mozilla XML User Interface Language (XUL).

References

- [1] Matthias Burger, Klaus Juenemann, and Thomas Koenig. *RUnit: R Unit test framework*, 2007. R package version 0.4.17.
- [2] The ECMAScript Group. *Standard ECMA-357. ECMAScript for XML (E4X) Specification*, 2nd edition edition, 2005.
- [3] Jeffrey Horner. *brew: Templating Framework for Report Generation*, 2007. R package version 1.0-2.
- [4] Duncan Temple Lang. *XML: Tools for parsing and generating XML within R and S-Plus.*, 2007. R package version 1.93-2.