libOR – library of OR data sets

Kenneth Sörensen\textsuperscript{a}, Kris Ven\textsuperscript{a}, Marc Sevaux\textsuperscript{b}, and Jan Verelst\textsuperscript{a}

\textsuperscript{a}University of Antwerp, Faculty of Applied Economics
\textsuperscript{b}University of Valenciennes, CNRS, UMR 8530, LAMIV-H-SP

Abstract

In this paper, we propose libOR, a new library of OR data sets, that greatly extends the functionality of existing platforms. libOR is decentralised, i.e. all users can add problem definitions and data sets (problem instances). Data sets in libOR are in XML format and can be validated online. Moreover, libOR offers the possibility to upload results obtained by a specific algorithm on a data set. This facilitates the comparison of different algorithms. All this is done through an easy-to-use web interface.

1 Introduction

In operations research, the quality of algorithms and heuristics is tested by applying them to a set of standard test problems. These test problems are distributed electronically, through e-mail or—most often—by putting them on a web site. The best-known of such web sites is the OR library\textsuperscript{1}. Originally described in [1], this web site has become the standard platform for distribution of data sets in the operational research community. For many years, the OR community has benefited from the centralised location of test problems that can be used for the evaluation of the performance of solution algorithms. However, the current OR Library lacks several interesting features:

- All data sets are in flat text format, very little semantic information is encapsulated. This makes it particularly hard to understand or even read the files and makes checking the correctness of a file virtually impossible.

\textsuperscript{1}http://www.ms.ic.ac.uk/info.html

\textsuperscript{2}Correspondence to: Kenneth Sörensen, University of Antwerp, Faculty of Applied Economics, Prinsstraat 13, B-2000 Antwerp
All updating of the library is done in a centralised way. This makes the library slow to respond to current developments.

The results of different algorithms on the various data sets are not included in the library. Researchers developing new algorithms are therefore forced to survey the entire literature to be able to compare the performance of their algorithm with that of the existing algorithms.

This talk proposes libOR, a new library of OR data sets that fixes these problems.

2 libOR — library of OR data sets

2.1 Features

- Data sets are in XML format. For each problem, an XML schema of the corresponding data sets is provided. Data sets can now be easily validated and accessed by using standard XML parsers. Many XML parsers are currently available (e.g. Xerces\(^2\)). Most of them are available as callable libraries, that can be easily incorporated into algorithm code, thus removing the need to write specialised input code for each algorithm.

- Data sets can be entered into the library in a decentralised way. Every researcher that develops a set of data sets for a specific problem, can upload these data sets. The uploaded data sets are checked for validity by comparing them to the XML schema for the problem.

- New problems can be proposed by submitting an XML schema file for the corresponding data sets. Upon submission, a request for comments (RFC) is issued. When the new problem is accepted by the research community, the XML schema and data sets can be added.

- Solutions and objective function values for each data set are included in the library. Entering these data is also done in a decentralised way. With minimal effort, researchers can share the results of newly developed algorithms with the community. Comparison between different solution approaches is facilitated.

- Each problem category has a responsible that looks after the consistency and the correctness of the data.

2.2 Structure

libOR is structured as shown in figure 1. Below is a description of the parts of libOR.

\(^2\)http://xml.apache.org
Problem category: A general container for several related problems (e.g. routing, scheduling, etc.). A problem category has a responsible that ensures the correctness and consistency of the data sets and results in libOR.

Problem: A description (including an XML schema file) of a specific OR-related problem, e.g. the vehicle routing problem with time windows (VRPTW).

Instance collection: A container for several instances of the same problem, that have been obtained e.g. by the same author and should be considered as a whole.

Instance: A specific data set in XML for a single problem instance.

Result collection: A container for a collection of results corresponding to an instance collection.

Result: A specific value for the result variables of a single problem instance.

Source: A description of the origin of a problem, an instance collection or a result collection, e.g. a paper or website.

2.3 Open Source

When developing libOR, we opted for using Open Source Software (OSS) as much as possible. Hence, the base system is a LAMP (Linux, Apache, MySQL, PHP) server, chosen for its high performance and reliability. To allow for on-line validation of XML files, we selected the Xerces libraries and the included SAX parser. Since LATEX is an essential tool for OR researchers,
texvc (part of Mediawiki\textsuperscript{3}) was integrated. This tool can convert \LaTeX\ formulas to a proper representation in HTML.

Our preference for OSS made us decide to release the code base of libOR under the terms of the GNU GPL\textsuperscript{4}. The project was named “ORLi” and will be available on the libOR website. By making ORLi Open Source, we want to make libOR as free as possible and invite the community to take actively part in the further development of libOR.

3 Conclusions

In this talk, we have presented libOR, a new library of OR data sets. libOR proposes several features that should increase the ease and correctness with which OR data sets are shared between researchers. Moreover, libOR allows researchers to enter results, which allows comparison of existing techniques without requiring a complete literature survey.

The success of libOR depends on the appreciation and cooperation of the research community. At this point, several interested persons have volunteered to be “responsible” for a certain problem category and are now in the process of building XML schema files for existing problems and converting existing data sets to XML formats. This should give libOR a kickstart and increase the acceptance with the research community.

References


\textsuperscript{3}http://wikipedia.sourceforge.net/
\textsuperscript{4}http://www.gnu.org/licenses/gpl.txt