

applications that must deal with large documents? Keep in mind that large here means too big to fit into memory, which is not so big if this format must be usable even for low-end desktop applications.

The current solutions in this space follow the so-called SAX (simple API for XML) approach: instead of building the whole tree of a document in one fell swoop and keeping it there for further processing, go step by step. A SAX parser reads the document and, instead of keeping it all in memory, generates an event every time it finds something worthwhile. The parser then passes the event to event handlers that interact with the application. The something worthwhile can be XML document-type definitions, errors or elements of the actual content. A good starting point for SAX-based programming is the SAX Project. SAX2 already is supported in Java through JAXP and in Perl through the Orchard Project, which is quite stable, not to mention fast and lightweight, as far as SAX and XML processing are concerned.

Conclusion

All the research done for this article confirmed one of my first impressions: so far, the free software/open-source software approach to guarantee information interchange has been to develop cross-platform applications, which are difficult to maintain and optimize for each target environment. Now it looks like we are starting to do the right thing, which is to define truly Free, standard, toolkit-independent, cross-platform formats that leave everyone free to create any possible front end to read and write them.

Acknowledgements

Thanks above all to Gary Edwards and David Faure for all the material and explanations. Pierre Souchay (kfile-plugin-ooo) and the AbiWord developers also were very helpful.

Marco Fioretti is a hardware systems engineer interested in free software both as an EDA platform and, as the current leader of the RULE Project, as an efficient desktop. Marco lives with his family in Rome, Italy.



Agata: www.agata.org.br



Agata is a PHP application that runs directly in PHP, not in a Web server. Why? I have no idea, but it works. And, if you need something that resembles (so I'm told) Crystal Reports

in Windows, take a look at Agata. It can be used against most SQL databases, although I tested it only against PostgreSQL. Agata is easy to install and run and is extremely flexible. It is also pretty and does graphs. Requires: PHP and php-gtk.

—DAVID A. BANDEL

RESOURCES

AbiWord: www.abisource.com

CPAN: www.cpan.org

EU-IDA: europa.eu.int/ISPO/ida

Expat: expat.sourceforge.net

kfile-plugin-ooo: bad.sheep.free.fr/kfile-plugin-ooo.html

KOffice: koffice.kde.org

KOffice DTD: www.koffice.org/DTD/kword-1.2.dtd

Libxml: xmlsoft.org

OASIS Office File Format TC: www.oasis-open.org/committees/tc_home.php?wg_abbrev=office

OASIS Web Site: <http://www.oasis-open.org>

OoO2sDbk: www.chez.com/ebellot/ooo2sdbk

ooo2txt: ooo2txt.free.fr

OpenOffice.org: www.openoffice.org

Orchard: orchard.sourceforge.net

QDom: doc.trolltech.com/3.1/xml-tools.html

RTF2XML: www.xmeta.com/omlette/rtf2xml

SAXEcho: xml.openoffice.org/saxecho

SAX Project: www.saxproject.org

SIAG, O3read, o3totxt, o3tohtml: siag.nu

soffice2html: hoopajoo.net/projects/soffice2html.html

Stop Word Attachments: www.gnu.org/philosophy/no-word-attachments.html

TclXML: tclxml.sourceforge.net

Writer2LaTeX: www.hj-gym.dk/~hj/writer2latex

Xerces: xml.apache.org/xerces-j

XMerge: xml.openoffice.org/xmerge

XMLsec: www.aleksey.com/xmlsec