BIOINFORMATICS APPLICATIONS NOTE

Vol. 00 no. 00 2008
Pages 1–3

An embedding feature of LyX

Bo Peng (ben.bob@gmail.com)

LyX Mechanics <lyx-devel@lists.lyx.org>

Received on June, 2008

The LyX Team

ABSTRACT

Unlike Word or OpenOffice, LyX/LaTeX users work mostly with external files. There are some advantages of using external files, but when it comes to sharing documents between co-authors, keeping track of all external files can be a hassle. This article describes an embedding feature of LyX that allows users embed part or all external files with a LyX document.

The design goal

The goal of this feature is allow easy sharing of LyX documents. Briefly,

- We need a LyX document format that bundles all needed external files.
- This document should be opened, viewed, compiled and edited directly, on any operating systems that LyX supports.
- The traditional way of handling of external files would better be preserved. This implies continuous use of insets with external files, and the ability to unbundle embedded files.

Embed, unembed, bundle and unbundle Embed / unembed and bundle / unbundle are different operations.

- **embed** Add an external file to a LyX document so that changes of the external version does not affect document output.
- **unembed** Extract an embedded file to an external file. The inset now refers to this external file.
- **bundle** Place external files along with a LyX document so that they can be sent together.
- **unbundle** Separate external files from a LyX document.
- **link** A map between internal file and its external source.

In summary, embed/unembed are inset level operations and bundle/unbundle are buffer level operations. It has been clear that

- **It is difficult to bundle/unbundle in a platform indepdent way**. For example, when `c:\figures\a.png` is inserted to a .lyx file under windows, it can not be directly unbundled under Linux. We should either disallow such files, or change it to another name during unbundling.
- In-place unbundling is unsafe. In-place unbundling refers to an operation that extract all bundled files to their original, potentially out of the document directory, locations. Such an operation can lead to security problem. Therefore, **unbundling should be performed in an empty directory**.
- File links can lead to security problems. Keeping file content and its original location in a LyX document can release private information on how external files are organized on the author's machine. **It should be possible to remove file link information, or encrypt them**.

A bundle/unbundle solution?

It has been obvious to many LyX developers that we need a bundle/unbundle solution. That is to say, we bundle all external files with a LyX document so that they can be sent to a co-author. This document would then be unbundled before editing. An seemingly obvious advantage of this approach is that LyX would always work with *external* files so little modification is needed to the LyX core.

Due to implementation difficulties, there has been no solid proposal that implements this method. Such a proposal is likely to suffer from the following problems:

- If all insets have to be embedded or unembedded, inserting an external file will embed this file in a bundle-editing mode under which all files are required to be embedded. This makes it different, if not more difficult, to work with external files in such a mode.
- A document may be modified just to be opened (unbundled) under another operating system. Keeping the same document and external files across different platforms is very complicated.

An embed/unembed solution

Inset-level embed/unembed is a clever way to avoid the mentioned problems of the bundle/unbundle solution. The basic idea is that a file can be embedded into a LyX document. If we embed all external files, the document is effectively bundled, if we unembed all external files, the document is effectively unbundled. However, there are two important differences:

- If we allow inset level embed/unembed. There is no need for a special bundled editing mode. External files can be used as before. There is no change to how users work with external files.
- More importantly, individual embedding allows partial unbundling. That is to say, if a file can not be unembedded on another platform, it can be kept embedded in the document and it will not cause any trouble. This partial in-place unbundle

is much easier to use than full-unbundling, which usually involves unbundling to an empty directory, close lyx, copy files back, and open the unbundled document.

Other key features of this solution include:

- Embedded files are invisible to users. They can have a link to its original external file, or be anonymous if such information is removed.
- Customized layout (.layout), class (.cls, .sty), bibtex style (.bst). and other files can be embedded. This allows the creation of a completely self-contained document.
- The embedded files are written directly in the LyX document encoded in base64. The existing compression feature can be used to compress this file.

Inset level features:

1. Inset context menu **Embedded** can be used to show the embedding status of an inset, and turn on/off embedding. This is a on/off switch for graphics, external and include inset, where unembed is only allowed when an external file is available and identical to the embedded version. Because a bibtex inset can have more than one embedded or external files, the context menu becomes 'Embed all bib files' and 'Unembed all bib files'.
2. Inset context menu **Save embedded file as**. Save the embedded file to an external file.
3. Inset context menu **Update from external file**. Update the embedded file from an external file.

Note that the external file used in operation 2 and 3 becomes the new 'external fle' of the embedded file. Un-embed can be then be used to make an inset refer to this external file.

Buffer level features:

1. **Document -> Embedded Files -> Embed all files**. This will embed all embeddable files in a buffer.
2. **Document -> Embedded Files -> Make embedded files anonymous**. Remove the original filenames of embedded files because the original name may contain sensible machine-specific information.
3. **Document -> Settings -> Embedded Files**. Arbitrary files such as layout and class files can be embedded through this dialog.
4. **Document -> Embedded Files -> Un-embed files with external copies**. This operation will un-embed all insets with an existing and identical external file. This feature can be used to in-place unbundle a file if only a few external files are changed in the embedded .lyx file.
5. **File -> Export -> LyX with extracted embedded files**. This operation will create a new directory filename.ext_lyx under the document directory, extract all embedded files with known original names in a way that the original external file structure can be preserved. For example, filename.lyx may be a level down that directory with an inset referring to ../figures/-figure.png. A Python script is used to allow such unbundling to be done outside of lyx.

Note that the export operation is not recommended because the exported .lyx file may not be compilable due to OS differences (e.g. if c:\blah is extracted as $DOC_DIR\filename.extracted\c\blah).

Expected workflow:

1. When a user needed to send a document to his co-author, he chose 'Document->Embedded Files->Embed all' and embed all the external files. He could optionally embed .cls, .layout, .bst files to make his document self-contained.
2. When a co-author received this document, he could view, edit and compile this document directly. When he finished editing the document, he could choose 'Embed all' to embed new external files.
3. The original author received the document and chose 'Document->Embedded Files->Unembed files with external copies'. All embedded files are unembedded except for the ones that have been modified or added by his co-author.
4. When an embedded file needs to be modified, a user can 'unembed' (or 'save as' then 'unembed' if there is no existing external file) to make this inset external, or 'update from external file' to update directly from an external file.
5. A user can extract all embedded files from a .lyx document using an export operation. Due to OS differences, the extracted file may not be compilable.
6. Optionally, a user can anonymise all embedded files to avoid saving machine-specific path information in the .lyx file. These embedded files will not be extracted during export.

DEMO INSETS

This articule uses an embedded layout, a customized class file with a bunch of .sty files, and a customized .bst file. They are accessible through the "Document->Settings->Embedded Files".

Here is an embedded graphics inset $\frac{a}{b}$. If it is in embedded status, a small pin will be shown at the left top corner of the figure.

An embedded program listing:

Listing 1. scripts/dir_copy.py

```
import shutil, sys, os

fromDir = sys.argv[1]
toDir = sys.argv[2]

if os.path.isdir(toDir):
    shutil.rmtree(toDir)

shutil.copytree(fromDir, toDir)
```

And some bib reference from two embedded bib files. (Abdallah *et al.*, 2003; Abdulrazzaq *et al.*, 1997).

Implementation

The feature will not be part of LyX 1.6.0 due to discrepancies of opinions between developers. I have nevertheness implemented my proposal and created a patch for LyX 1.6.0 beta 2. If you are intersted, you can apply the patch http://www.lyx/org/~bpeng/embedding/embed.diff. Please also run 'svn mv lib/images/attic/pin.png lib/image' before you compile.

To test the patch, you can

- Insert graphics, child documents etc, embed and unembed.
- Open an existing document, embed all insets and unembed them all.
- Export the document to LyX Document with extracted embedded files.
- Move these documents to another directory and try to unembed all insets.
- Try to update or unembed insets using 'save embedded file as', 'update embedded file from' and 'unembed' features.

Enjoy!

REFERENCES

Abdallah, J. M., Goffinet, B., Cierco-Ayrolles, C., and Perez-Enciso, M. (2003). Linkage disequilibrium fine mapping of quantitative trait loci: a simulation study. *Genet Sel Evol*, **35**, 513–532.

Abdulrazzaq, Y., Bener, A., Al-Gazali, L., Al-Khayat, A., Micallef, R., and Gaber, T. (1997). A study of possible deleterious effects of consanguinity. *Clin Genet*, **51**, 167–173.