

Hacking HSSF

by Glen Stampoulzis, Andrew Oliver

1. Where Can I Find Documentation on Feature X

You might find the 'Excel 97 Developer's Kit' (out of print, Microsoft Press, no restrictive covenants, available on Amazon.com) helpful for understanding the file format.

Also useful is the [open office XLS spec](#). We are collaborating with the maintainer of the spec so if you think you can add something to their document just send through your changes.

2. Help, I Can't Find Feature X Documented Anywhere

1. Look at OpenOffice.org or Gnumeric sources if its implemented there.
2. Use `org.apache.poi.hssf.dev.BiffViewer` to view the structure of the file. Experiment by adding one criteria entry at a time. See what it does to the structure, infer behavior and structure from it. Using the unix diff command (or get cygwin from www.cygwin.com for windows) you can figure out a lot very quickly. Unimplemented records show up as 'UNKNOWN' and prints a hex dump.

3. Low-level Record Generation

Low level records can be time consuming to created. We created a record generator to help generate some of the simpler tasks.

We use XML descriptors to generate the Java code (which sure beats the heck out of the PERL scripts originally used ;-)) for low level records. The generator is kinda alpha-ish right now and could use some enhancement, so you may find that to be about 1/2 of the work. Notice this is in `org.apache.poi.hssf.record.definitions`.

4. Important Notice

One thing to note: If you are making a large code contribution we need to ensure any participants in this process have never signed a "Non Disclosure Agreement" with Microsoft, and have not received any information covered by such an agreement. If they have they'll not be able to participate in the POI project. For large contributions we may ask you to sign an agreement.

5. What Can I Work On?

Check our [todo list](#) or simply look for missing functionality. Start small and work your way up.

6. What Else Should I Know?

Make sure you [read the contributing section](#) as it contains more generation information about contributing to POI in general.