Formula Support

by Avik Sengupta

1. Introduction

This document describes the current state of formula support in POI. The information in this document currently applies to the 3.5 version of POI. Since this area is a work in progress, this document will be updated with new features as and when they are added.

2. The basics

In org.apache.poi.hssf.usermodel.HSSFCell **setCellFormula("formulaString")** is used to add a formula to sheet and **getCellFormula**() is used to retrieve the string representation of a formula.

We aim to support the complete excel grammar for formulas. Thus, the string that you pass in to the *setCellFormula* call should be what you expect to type into excel. Also, note that you should NOT add a "=" to the front of the string.

3. Supported Features

- References: single cell & area, 2D & 3D, relative & absolute
- Literals: Number, text, boolean, error and array
- Operators: arithmetic and logical, some region operators
- Built-in functions: over 350 recognised, 280 evaluatable
- Add-in functions: 3 from Analysis Toolpack

4. Not yet supported

- Manipulating array/table formulas (In Excel, formulas that look like "{=...}" as opposed to "=...")
- Region operators: union, intersection
- Parsing of previously uncalled add-in functions
- Preservation of whitespace in formulas (when POI manipulates them)

5. Internals

Formulas in Excel are stored as sequences of tokens in Reverse Polish Notation order. The open office XLS spec is the best documentation you will find for the format.

The tokens used by excel are modelled as individual *Ptg classes in the **org.apache.poi.hssf.record.formula** package.

The task of parsing a formula string into an array of RPN ordered tokens is done by the **org.apache.poi.ss.formula.FormulaParser** class. This class implements a hand written recursive descent parser.

Formula tokens in Excel are stored in one of three possible *operand classes* : Reference, Value and Array. Based on the location of a token, its class can change in complicated and undocumented ways. While we have support for most cases, we are not sure if we have covered all bases (since there is no documentation for this area.) We would therefore like you to report any occurrence of #VALUE! in a cell upon opening a POI generated workbook in excel. (Check that typing the formula into Excel directly gives a valid result.)

Check out the javadocs for details.