Apache POI - HPSF - Java API to Handle Microsoft Format Document Properties

Overview

by Rainer Klute

1. Overview

Microsoft applications like "Word", "Excel" or "Powerpoint" let the user describe his document by properties like "title", "category" and so on. The application itself adds further information: last author, creation date etc. These document properties are stored in so-called **property set streams**. A property set stream is a separate document within a POI filesystem. We'll call property set streams mostly just "property sets". HPSF is POI's pure-Java implementation to read and write property sets.

The HPSF HOWTO describes what a Java application should do to read a property set using HPSF, how to retrieve the information it needs, and how to write properties into the document.

HPSF supports OLE2 property set streams in general, and is not limited to the special case of document properties in the Microsoft Office files mentioned above. The HPSF description describes the internal structure of property set streams. A separate document explains the internal of thumbnail images.