

PowerPoint

Extracting Text from PowerPoint format

Here are different ppt extraction code. No guarantees, please modify list information if you test it.

Using Apache Tika: <http://tika.apache.org/>

Using POI HSLF: [Quick Guide](#) (see [<http://jakarta.apache.org/poi/hslf/quick-guide.html>] for details on text extraction)

From: poi-users: <http://www.mail-archive.com/poi-user@jakarta.apache.org/msg04809.html>

From: slide-dev: <http://www.mail-archive.com/slide-dev@jakarta.apache.org/msg10445.html>

From: <http://nagoya.apache.org/eyebrowse/ReadMsg?listName=poi-dev@jakarta.apache.org&msgNo=4326>

Here is some sample code that works with 'some* ppt formats. It's basically an implementation of a *POIFSReader**Listener. There are no guarantees on how well it works - it is known to ignore unicode text records for starters. It requires POI libraries.

[PPT2Text.java](#)