

Apache POI - Encryption support

by Maxim Valyanskiy

1. Overview

Apache POI contains support for reading few variants of encrypted office files:

- XLS - RC4 Encryption
- XML-based formats (XLSX, DOCX and etc) - AES and Agile Encryption

Some "write-protected" files are encrypted with build-in password, POI can read that files too.

2. XLS

When HSSF receive encrypted file, it tries to decode it with MSOffice build-in password. Use static method `setCurrentUserPassword(String password)` of `org.apache.poi.hssf.record.crypto.Biff8EncryptionKey` to set password. It sets thread local variable. Do not forget to reset it to null after text extraction.

3. XML-based formats

XML-based formats are stored in OLE-package stream "EncryptedPackage". Use `org.apache.poi.poifs.crypt.Decryptor` to decode file:

```
EncryptionInfo info = new EncryptionInfo(filesystem);
Decryptor d = Decryptor.getInstance(info);

try {
    if (!d.verifyPassword(password)) {
        throw new RuntimeException("Unable to process: document is encrypted");
    }

    InputStream dataStream = d.getDataStream(filesystem);

    // parse dataStream

} catch (GeneralSecurityException ex) {
    throw new RuntimeException("Unable to process encrypted document", ex);
}
```

If you want to read file encrypted with build-in password, use
Decryptor.DEFAULT_PASSWORD.