

Practical Office Automation

Dentro il Formato di File OpenOffice.org

Jacob Sparre Andersen

sparre@crs4.it



GULCh

Gruppo Utenti Linux Cagliari h...?

Ouverture

- **Subject:**

Questo talk e' sull'uso e l'estrazione dei meta-data dai file OpenOffice.org/OpenDocument

- **Audience:**

System administrators, system programmers and information system decision makers.

Discutero' su cosa voi potete (dire ai vostri programmatori) di fare, se i vostri documenti sono in un formato aperto. Una volta discusso cosa voi *potete* fare, daro' alcuni esempi di *come* farlo con tools standard di Linux.

Le domande sono ben accolte *in ogni momento* durante il talk.

Rassegna (1)

- *Cosa puoi fare, se i tuoi documenti sono in formato aperto.*
- Contenuto di un file OpenOffice.org.
- Indicizzazione di documenti OpenOffice.org.
- Prevenzione di fuga di documenti riservati attraverso firewall.

Formati aperti

I requisiti minimi per uno standard aperto sono che il formato del documento sia completamente descritto in documenti pubblicamente accessibili, [...] e che il formato del documento possa essere implementato in programmi senza alcuna restrizione, libero da brevetti e senza implicazioni legali.

<http://europa.eu.int/idabc/servlets/Doc?id=17982>

Benefici dell'uso dei formati aperti

- Nessun legame con un singolo fornitore di software.
- Minor prezzo sul software off-the-shelf.
- Liberta' di (fare vostri programmatori) implementare tools in casa.
- It is more likely that you can find Open Source programs which already solve your problems.

Idee per tools "fatti in casa"

Since the OpenOffice.org/OpenDocument file format is open, you are free to (make your programmers) write special tools for manipulating the files:

- Estrazione di titoli e parole chiave per l'indicizzazione automatica dei documenti
- Prevenire che documenti riservati per la propria società escano attraverso il firewall aziendale.
- Warning authors about lacking project codes in documents.
- ...

Only your imagination (and your ability to explain your ideas) sets limits.

Rassegna (2)

- Cosa puoi fare, se i tuoi documenti sono in formato aperto.
- *Contenuto di un file OpenOffice.org.*
- Indicizzazione di documenti OpenOffice.org.
- Prevenzione di fuga di documenti riservati attraverso firewall.

Contenuto di un file OpenOffice.org (1)

```
% unzip -l skriv-og-slet.sxw
```

Length	Date	Time	Name
-----	----	----	----
30	11-14-05	11:40	mimetype
1958	11-14-05	11:40	content.xml
5979	11-14-05	11:40	styles.xml
1282	11-14-05	11:40	meta.xml
6280	11-14-05	11:40	settings.xml
752	11-14-05	11:40	META-INF/manifest.xml
-----			-----
16281			6 files

Contenuto di un file OpenOffice.org (2)

```
% unzip -ap skriv-og-slet.sxw meta.xml \  
> | sed 's/></>\n</g'  
<?xml version="1.0" encoding="UTF-8"?>  
<!DOCTYPE office:document-meta PUBLIC "-//OpenOf  
<office:document-meta xmlns:office="http://openc  
<office:meta>  
<meta:generator>OpenOffice.org 1.1.4 (Unix)</met  
<!--645 (Build:8824)-->  
<dc:title>Writes and deletions</dc:title>  
<meta:creation-date>2005-11-14T12:31:10</meta:cr  
<dc:date>2005-11-14T12:40:48</dc:date>  
<meta:keywords>
```

Contenuto di un file OpenOffice.org (3)

```
> | sed 's/></>\n</g' \ > | grep '<meta:keyword>  
<meta:keyword>00o</meta:keyword> <meta:keyword>fi  
<meta:keyword>demonstration</meta:keyword>  
<meta:keyword>changes</meta:keyword>
```

Contenuto di un file OpenOffice.org (4)

```
% unzip -ap skriv-og-slet.sxw meta.xml \  
> | sed 's/></>\n</g' \  
> | grep '<dc:title>'  
<dc:title>Writes and deletions</dc:title>  
%
```

Contenuto di un file OpenOffice.org (5)

```
% unzip -ap skriv-og-slet.sxw content.xml \  
> | sed 's/></>\n</g' \  
> | grep '<text:tracked-changes>'  
<text:tracked-changes>  
%
```

Rassegna (3)

- Cosa puoi fare, se i tuoi documenti sono in formato aperto.
- Contenuto di un file OpenOffice.org.
- *Indicizzazione di documenti OpenOffice.org.*
- Prevenzione di fuga di documenti riservati attraverso firewall.

Indicizzazione di documenti OpenOffice.org

Practical demonstration of indexing of OpenOffice.org documents.

Rassegna (4)

- Cosa puoi fare, se i tuoi documenti sono in formato aperto.
- Contenuto di un file OpenOffice.org.
- Indicizzazione di documenti OpenOffice.org.
- *Prevenzione di fuga di documenti riservati attraverso firewall.*

Checking for document histories

Practical demonstration of checking OpenOffice.org documents for change information.

Più informazione

- A commented command history from the practical demonstrations will be published on <http://edb.jacob-sparre.dk/foredrag/00o/> after the talk.
- Write me at sparre@crs4.it if you have questions related to the talk.

Fine.