

Open Access Media Importer

Wissenswert 2011

Idee: Daniel Mietchen

Wikipedia hat einen
Mangel an Multimedia

Open Access

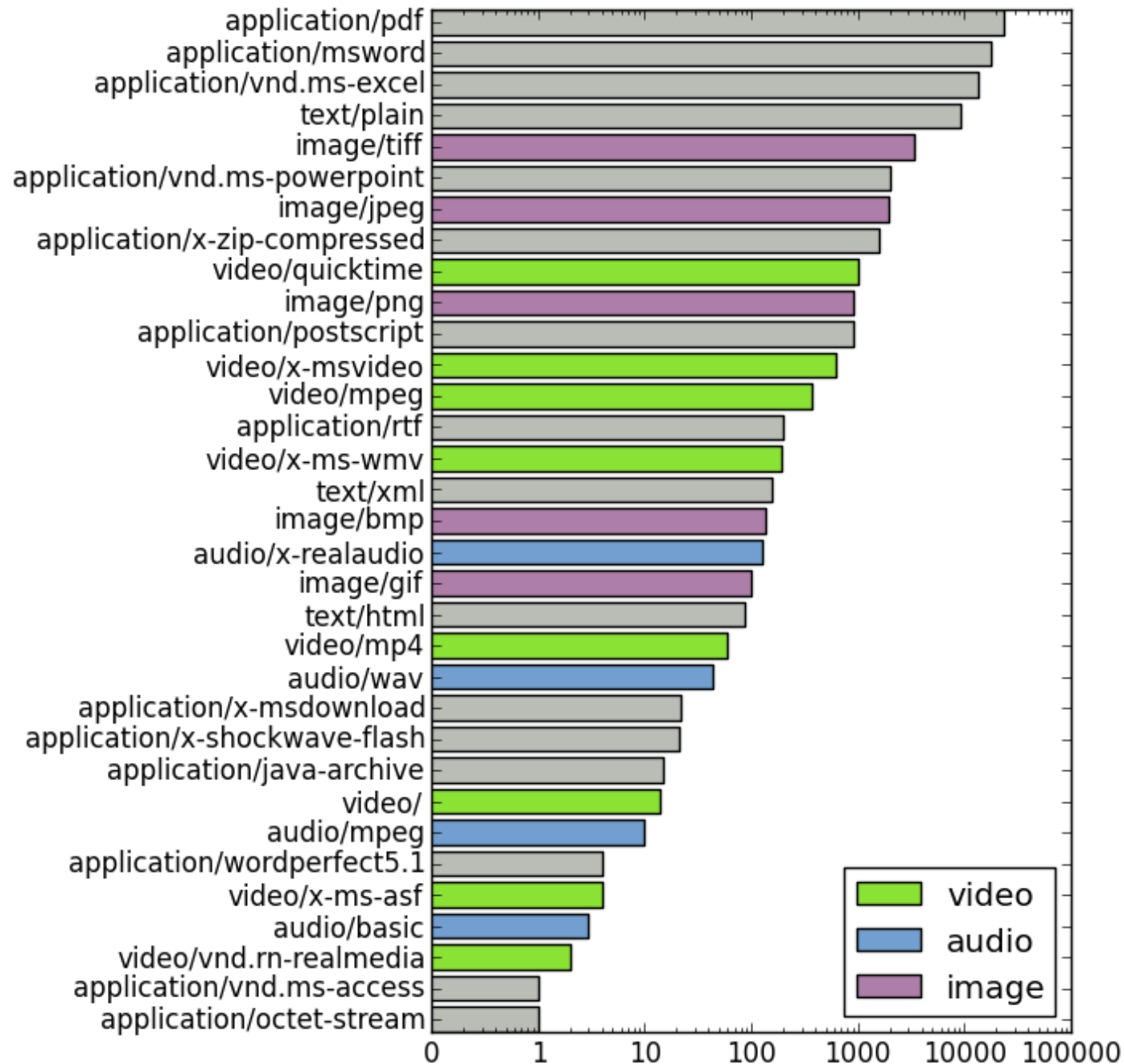
Artikel mit Audio / Video

Medien → Archive

Medien → Commons

PubMed Central

MIME Types of Supplementary Materials under Free Licenses



Implementation

Vorbild: *apt-get*

modularer Aufbau

oa-get

oa-cache

oa-put

oa-get

Download

oa-get

- Metadaten holen
- Medien holen

oa-cache
Processing

oa-cache

- Metadaten ermitteln
- Format-Umwandlung
- Datenbank-Zeugs
- ...

oa-put
Upload

oa-put

- Template
- Duplikatcheck
- Medien hochladen

Datenquellen als Module

- *pmc*
- *pmc_doi*
- *pmc_xml*

Beispiel-Workflow

oa-get download-metadata pmc
oa-cache find-media pmc
oa-get update-mimetypes pmc
oa-get download-media pmc
oa-cache convert-media pmc
oa-put upload-media pmc

Freie Software – GPL v3+

Alle Abhängigkeiten in *Debian*!

- Python 2.7
- GStreamer
- Mutagen
- ...

Blog: <<http://wir.okfn.org/>>

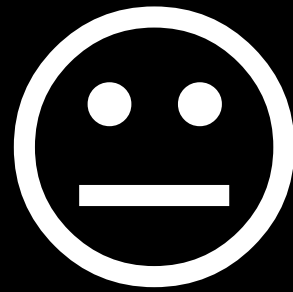
Code: <<https://github.com/erlehm/ann-open-access-media-importer>>

Probleme

XML uneinheitlich

<license>

<copyright-statement>



```
<license xlink:href="http://creativecommons.org/licenses/by/3.0/" />
```


<license>

This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

</license>

<license>

This is an open access paper distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

</license>

<copyright-statement>

This is an open-access article distributed under the terms of the Creative Commons Public Domain declaration, which stipulates that, once placed in the public domain, this work may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose.

<copyright-statement>



<license>

This work is licensed under a Creative Commons Attribution 3.0 License
(by-nc 3.0) Licensee PAGEPress, Italy

</license>

<license>

This work is licensed under a Creative Commons Attribution 3.0 License (by-nc 3.0). Licensee PAGE Press, Italy

</license>



```
<license>
```

```
Open access articles can be viewed online without a subscription.
```

```
</license>
```



```
<license>
```

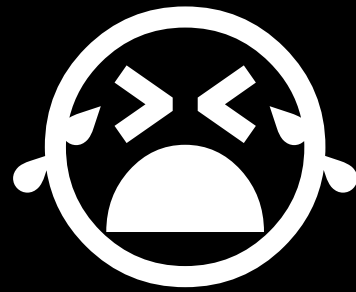
```
creative commons
```

```
</license>
```

```
<license>
```

```
open access
```

```
</license>
```



MIME-Type

	text/xml	video/x-ms-wmv	application/x-zip-compressed	text/plain	application/octet-stream	application/pdf
video/x-ms-asf		5				
application/vnd.ms-powerpoint				3		
application/zip			3			1
text/plain					1	
application/vnd.ms-excel			2			
application/xml	1					
application/pdf				8		
application/msword				2		

min. 1% falsche MIME-Types

oa-get update-mimetypes

One more thing ...

Quicktime-Videos



Amazing!

F-~~AAV~~-LRAT



Control

Diskussion

- Dateinamen?
- Kategorien?
- Workflow?
- Toolserver?