

BibTeX Bibliography Index Maker: Notes

Ramon Xuriguera

1 BibTeX

Aspectes del format BibTeX a tenir en compte:

- Com podem distingir entre diferents tipus d'entrada (article, book, inproceedings, etc.) a partir del fitxer?
- Format dels noms. Un nom consisteix de diferents parts: First, von, Last, Jr. El token *von* o *de la* cal posar-los en minúscules. Per tal que el nom es reconegui, cal que tingui el format: von Last, Jr, First. D'aquesta manera, si hi ha més d'un cognom no passa res.
- Caràcters Unicode entre claus per poder ser utilitzats correctament amb l'estil *alpha*. Per exemple: `Jos{\'\{e\}}`
- Per prevenir que BibTeX canviï un text a minúscules, cal posar el text entre claus.
- Si hi ha massa autors, truncar la llista amb *et al.*
- Utilitzar abreviatures de tres lletres per als mesos
- Utilitzar el camp **key** per a organitzacions amb un nom llarg, de manera que s'utilitzin les inicials de l'organització al fer una cita.

2 PDF

Aspectes a considerar:

- Caràcters Unicode
- Glyphs com ara *fi* corresponen a més d'una lletra
- Les llibreries extreuen el text per línies
- Podem obtenir les següents dades del fitxer sense haver-ne d'extreure el text: número de pàgines, títol, autor, assumpte i paraules clau. Si aquests camps no s'han omplert al generar el fitxer, estaran en blanc.

2.1 Text extraction

2.1.1 PDFBox

Llibreria escrita en Java i publicada sota la llicència *Apache License v2.0*. Actualment es troba a la incubadora d'Apache.

2.1.2 PDFMiner

Programes desenvolupats en Python per extreure i analitzar dades de documents PDF.

2.1.3 jPod

3 Articles Db

DBLP

DBLP proporciona una API molt simple per cercar autors autors i obtenir les seves publicacions, però no al revés. Al cercar per autor, es mostren les claus de cada publicació, però no el títol.

Portal ACM

Google Scholar

Enllaç directe a la cita en format BibTeX. Actualment no proporcionen cap API per obtenir-ho. Caldria fer-ho amb `wget`.

CiteSeerX

Arxiv.org

<http://arxiv.org/help/api/index>

4 Software existent

http://en.wikipedia.org/wiki/Comparison_of_reference_management_software

JabRef

Java, llicència: GPL.

És possible crear plug-ins amb *Java Plug-in Framework*.

<http://jabref.sourceforge.net/>

Mendeley

Propietari, no accepta plug-ins.

<http://www.mendeley.com/>

BibDesk

Objective C, BSD.

<http://bibdesk.sourceforge.net/>

5 Llenguatge

- Java
- Python
- Java + Python amb Jython
- C# i IronPython