

BibTeX Bibliography Index Maker: Meeting Notes

Ramon Xuriguera

01-02-2010

1 Extracció d'informació

Com fer-ho? Com començar?

1. Wrappers creats a mà per cadascun dels diferents llocs web dels quals vulguem extreure informació
2. Wrapper induction
3. Ontologia + ML per omplir els camps

De moment ens quedem amb la opció 1 (hardcode)

2 ToS de Google i Google Scholar

1.4 Appropriate Conduct. **You shall not**, and shall not allow any third party to: (a) edit, modify, truncate, filter or change the order of the information contained in any Results (either individually or collectively), including, without limitation, by way of commingling Results with non-Google provided search results or advertising; (b) frame any Results or Results Page; (c) display any Results in pop-up, pop-under, exit windows, expanding buttons, or animation; (d) display any Results to any third parties other than End Users; (e) minimize, remove or otherwise inhibit the full and complete display of any Results Page (including any Results); **(f) directly or indirectly access, launch and/or activate the Service through or from, or otherwise incorporate the Service in, any Client Application, Web site or other means other than the Site, and then only to the extent expressly permitted herein;** (g) transfer, sell, lease, syndicate, sub-syndicate, lend, or use for co-branding, timesharing, service bureau or other unauthorized purposes any Service or access thereto (including, but not limited to Results, or any part, copy or derivative thereof); (h) enter into any arrangement or agreement under which any third party pays You fees, You pay any third party fees, or either shares in any revenue payments and/or royalties for any Results; **(i) directly or indirectly generate queries**, or impressions of or clicks on Results, **through any automated**, deceptive, fraudulent **or other invalid means** (including, but not limited to, click spam, robots, macro programs, and Internet agents); **(j) modify, adapt, translate, prepare derivative works from**, decompile, **reverse engineer**, disassemble or otherwise attempt to derive source code from any Service or any other Google technology, content, data, routines, algorithms, methods, ideas design, user interface techniques, software, materials, and documentation; (k) remove, deface, obscure, or alter Google's copyright

notice, trademarks or other proprietary rights notices affixed to or provided as a part of any Service or any other Google technology, software, materials and documentation; (l) "crawl", "spider", index or in any non-transitory manner store or cache information obtained from the Service (including, but not limited to, Results, or any part, copy or derivative thereof); (m) create or attempt to create a substitute or similar service or product through use of or access to any of the Service or proprietary information related thereto; and/or (n) engage in any action or practice that reflects poorly on Google or otherwise disparages or devalues Google's reputation or goodwill. Further, the Site shall not contain any pornographic, hate-related or violent content or contain any other material, products or services that violate or encourage conduct that would violate any criminal laws, any other applicable laws, or any third party rights.

De moment, no ho tenim en compte. Ja ens ho mirarem millor a l'hora de distribuir la utilitat o plug-in de JabRef.

3 Llicència

L'eina `pdftotext` de *xPDF* utilitza la llicència GPL. Això implica que també haurem d'utilitzar GPL. De moment, l'altra llibreria utilitzada, `xgoogle`, usa la llicència MIT que és GLP-compatible.

Utilitzem GPL

4 Nom de l'aplicació i dels paquets

Quin nom hem de ficar a l'aplicació i als paquets. Fins ara he fet servir tres variacions del mateix nom de diferent longitud: Bibtex Bibliography Index Maker, BibtexIndexMaker, bimaker (pels paquets python).

Canvi de nom dels paquets, de *bimaker* a *bibim*.