

BIB_TE_X Bibliography Index Maker: Meeting Notes

Ramon Xuriguera

18-05-2010

1 Proves

Proves que podem realitzar:

- Generació de Wrappers
 - Comparar la qualitat dels *wrappers* generats segons el número d'exemples a partir dels quals s'ha obtingut.
Per un parell de pàgines web mirar com varien els wrappers generats segons el número de referències a partir dels quals s'obtenen. Provar amb 2, 3, 4, 5...
 - Comparar el número de camps d'informació que ofereix una pàgina respecte els wrappers que s'han pogut generar correctament (o gairebé)
- Cerca de referències Avaluar els resultats obtinguts al cercar referències segons els diferents paràmetres:
 - Mínim/Màxim de paraules de la consulta
 - Consultes inicials que se salten per tal de descartar resultats on només hi ha el títol
 - Màxim nombre de consultes que a provar
 - Número de resultats que fan que es consideri la cerca massa poc restrictiva (Too-many-results)

Anar canviant els valors dels paràmetres i comparar el número de resultats i la qualitat segons cada cercador (Scholar, Google, Bing i Yahoo)

- Extracció de referències
 - Establir uns quants directoris amb PDFs d'una pàgina en concret
 - Limitar resultats de la web a la pàgina a la qual pertanyen els documents
 - Extreure les referències
 - Comptabilitzar referències encerts i comparar entre les diferents pàgines

2 Guió Presentació

- Començar amb BD buida
- Importar unes quantes referències des d'un fitxer BIB_TE_X

- Generar wrappers a partir d'aquestes referències. Fer-ho per un parell o tres de pàgines web.
- Escollir un directori amb uns quants PDF i extreure les referències.
- Mostrar les referències obtingudes i modificar errors (resultats bons però no perfectes)
- Exportar les referències a un fitxer