Títol: **BibTEX Bibliography Index Maker**

Volum: **1/1**
Alumne: **Ramon Xuriguera Albareda**

Director/Ponent: **Marta Arias**
Departament: **LSI**
Data: **Primavera 2010**

DADES DEL PROJECTE

Títol del Projecte:

Nom de l'estudiant: Ramon Xuriguera Albareda

Titulació: Enginyeria Informàtica

Crèdits: 37,5

Director/Ponent: Marta Arias

Departament: LSI

MEMBRES DEL TRIBUNAL *(nom i signatura)*

President:

Vocal:

Secretari:

QUALIFICACIÓ

Qualificació numèrica:

Qualificació descriptiva:

Data:

Índex

Capítol 1

Introducció

1.1 Descripció

At vero eos et accusamus et iusto odio dignissimos ducimus qui blanditiis praesentium voluptatum deleniti atque corrupti quos dolores et quas molestias excepturi sint occaecati cupiditate non provident, similique sunt in culpa qui officia deserunt mollitia animi, id est laborum et dolorum fuga. Et harum quidem rerum facilis est et expedita distinctio [Jr06].

1.2 Treball Existent

Nam libero tempore, cum soluta nobis est eligendi optio cumque nihil impedit quo minus id quod maxime placeat facere possimus, omnis voluptas assumenda est, omnis dolor repellendus.

Capítol 2

Definició del Projecte

2.1 Context

Phasellus eu ante diam, eu euismod nunc. Vivamus non dolor sem. Sed id metus enim. Curabitur consectetur eleifend quam porta sagittis. Mauris sed augue fermentum leo pharetra posuere nec euismod risus. In dui elit, iaculis eget vestibulum eu, suscipit at purus. Mauris hendrerit condimentum velit, in facilisis dui consectetur non. Quisque tristique velit vitae enim posuere suscipit. Integer condimentum rutrum accumsan. Suspendisse bibendum urna eget orci aliquam faucibus congue urna consequat. Nam elementum, lectus a volutpat gravida, felis nibh faucibus nibh, id fringilla arcu purus sed orci.

2.2 El format BibTeX

Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo. Nemo enim ipsam voluptatem quia voluptas sit aspernatur aut odit aut fugit, sed quia consequuntur magni dolores eos qui ratione voluptatem sequi nesciunt.

2.3 Característiques

At vero eos et accusamus et iusto odio dignissimos ducimus qui blanditiis praesentium voluptatum deleniti atque corrupti quos dolores et quas molestias excepturi sint occaecati cupiditate non provident, similique sunt in culpa qui officia deserunt mollitia animi, id est laborum et dolorum fuga. Et harum quidem rerum facilis est et expedita distinctio.

2.4 Planificació Temporal

Vestibulum eu purus turpis. Quisque vel tortor urna. Sed vehicula dui vel mauris euismod consectetur. Phasellus imperdiet, erat sed mattis volutpat, orci urna pulvinar dui, non vulputate nisl augue in nulla. Curabitur tempus pharetra nisi et convallis. Morbi tincidunt lacus eu ligula rutrum sed sodales turpis suscipit. Quisque porta urna in enim aliquam congue. Quisque felis augue, rutrum eget posuere id, tincidunt vel nisl. Nullam fringilla ullamcorper lectus, quis faucibus nibh vulputate vitae. Nulla venenatis condimentum justo id aliquet. Aliquam vulputate

lorem ut diam blandit at aliquet diam viverra. Nunc leo nibh, adipiscing vel posuere eget, ultrices sollicitudin dolor. Cras id leo leo, eu sagittis libero. Quisque vitae erat lorem, non elementum libero. Sed vitae dolor tortor, varius semper diam. Praesent consequat consequat augue sed consequat. Duis ut libero neque, at porta eros. Nunc ullamcorper libero in velit ornare congue. Quisque condimentum consequat enim nec volutpat. Pellentesque eu arcu ante, in sagittis enim. Cras interdum sagittis risus, eget sodales ipsum vestibulum quis. Suspendisse ut libero nec metus sollicitudin tincidunt. Nunc id elit odio, et facilisis lorem. Suspendisse ut turpis non lacus ultricies convallis aliquam in nunc. Vestibulum et felis enim, in cursus sem. Sed mauris diam, malesuada eu convallis a, pellentesque ultrices purus. Quisque cursus fermentum ultrices. Nulla facilisi.

Morbi in erat nec neque adipiscing laoreet. Praesent lobortis justo eget lacus egestas nec auctor nisl dictum. Nulla eu gravida justo. Praesent volutpat tincidunt condimentum. Quisque eu erat libero. Nam imperdiet molestie accumsan. Nulla non lectus ligula. Vestibulum non mi nunc, et faucibus neque. Pellentesque enim lorem, pretium non adipiscing sit amet, pretium et lacus. Duis sagittis faucibus interdum. Nulla facilisi. Maecenas nisl leo, fermentum non tincidunt eu, dignissim vel ipsum. Pellentesque sodales tempor est, id egestas leo eleifend at. Maecenas fringilla rhoncus iaculis. Nulla facilisi.

Etiam nunc urna, rutrum eu lacinia a, gravida eu justo. Proin cursus iaculis velit, id placerat ipsum facilisis facilisis. Mauris molestie sodales imperdiet. Suspendisse lorem neque, feugiat non euismod et, fringilla eget tortor. Integer aliquet, ante id congue rhoncus, metus ligula varius purus, ac facilisis diam sem a justo. Fusce ante arcu, vestibulum eu dapibus nec, pharetra vitae arcu. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Sed condimentum quam a velit sodales condimentum. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc scelerisque molestie tortor, id vehicula lacus tincidunt at. Nam lacus turpis, elementum vitae suscipit id, lacinia eget dui. Donec vel augue erat, sit amet iaculis sem. Morbi eget lorem in orci vehicula bibendum eget ac mi. Sed molestie vehicula nisi, ac rhoncus mi aliquam vel. Duis vitae dui odio, quis ullamcorper felis. Sed purus quam, placerat id interdum elementum, convallis eget elit.

Phasellus suscipit lorem ut nisi luctus eget volutpat mauris consectetur. Ut elit libero, accumsan sed condimentum tempus, hendrerit id felis. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc venenatis pellentesque arcu, in sagittis orci ullamcorper eget. Integer eros lacus, adipiscing eu faucibus eget, pellentesque at augue. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris metus mauris, euismod non mollis sit amet, bibendum vel risus. Etiam ipsum nunc, iaculis vitae malesuada a, ullamcorper ut orci. Mauris quis consequat erat. Proin posuere consequat risus nec tempor. Nunc velit sapien, consectetur at condimentum at, ultrices ac nisl. Nam ullamcorper, dolor vitae faucibus porta, sapien sem lacinia ligula, non laoreet velit lacus eget purus.

Nulla commodo lorem scelerisque ipsum accumsan eget semper felis laoreet. Maecenas sit amet ipsum nec mauris congue interdum. Fusce quis aliquet nunc. Praesent risus diam, tincidunt sit amet pharetra id, aliquam sed sapien. Morbi massa dolor, convallis sit amet aliquet suscipit, convallis quis ante. Vivamus a aliquam nisi. Sed adipiscing varius nibh a mollis. Aenean non orci metus. Curabitur porta sagittis sapien a blandit. Nunc convallis imperdiet justo eget porttitor. Donec semper blandit eros et accumsan. Morbi ante magna, euismod ut vestibulum venenatis, gravida ut eros. Nullam sed elit nec nisl feugiat aliquam vitae in tellus. Donec lorem magna, volutpat in fringilla sed, pharetra nec dolor. Nam ultricies lorem sit amet massa vestibulum id cursus purus fermentum. Donec condimentum sapien in mauris faucibus convallis. Fusce porta viverra enim ut sodales. Sed consequat tincidunt mi sit amet dictum. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Sed ac augue at felis venenatis hendrerit sit amet vitae ligula.

Suspendisse aliquam consequat ornare. Fusce in lacus id metus fringilla lobortis nec et ipsum. Sed et turpis velit, sed faucibus lacus. Etiam et sem at odio consequat facilisis. Ut lorem leo, imperdiet a congue vitae, euismod non ipsum. Vestibulum pretium, tortor sed venenatis adipiscing, erat ipsum mattis erat, tristique lobortis eros lectus lacinia mi. Integer et orci libero. Mauris odio est, convallis convallis porttitor eu, blandit non sem. Praesent cursus erat vel nunc pulvinar porttitor. Duis tincidunt elementum mauris quis sodales. Donec euismod, justo quis malesuada vulputate, sem dui tincidunt justo, ut consectetur nulla velit in arcu. Nam viverra, odio eu hendrerit viverra, dolor ipsum ornare nunc, eu vestibulum arcu purus eget erat. Fusce sodales vehicula erat vitae consequat. Nulla fermentum feugiat orci eu tristique. In in fringilla neque. Morbi est libero, tempor eu hendrerit eget, tempor ac tortor. Cras leo turpis, mattis ut viverra at, vehicula vel enim. Maecenas pellentesque libero in risus porttitor in vulputate ligula commodo. Ut pharetra lobortis viverra.

Aenean euismod pulvinar dui at porttitor. Praesent nec velit tortor, ac aliquam elit. Ut id imperdiet justo. Cras vel nulla quis est tempus posuere nec rhoncus metus. Sed laoreet purus at ante tempus non sodales arcu varius. Curabitur auctor lacinia adipiscing. Nulla nec purus metus. Nullam ut metus vitae eros ullamcorper pretium et et justo. Ut vestibulum libero vel risus accumsan adipiscing. Morbi tempor facilisis ligula, sed mattis dolor bibendum eu. Etiam imperdiet interdum arcu, sed interdum mi semper nec. Ut a euismod libero. Pellentesque fringilla turpis id nisi tristique euismod. Pellentesque in felis a tellus laoreet viverra. Integer nibh ante, euismod id adipiscing sed, viverra in nisl. Maecenas sem lorem, lobortis nec adipiscing ut, consectetur sed turpis.

Proin orci metus, blandit fringilla viverra in, bibendum vel arcu. Vestibulum et enim id lorem ornare commodo. Integer ornare tempus metus, ac dictum nisi sagittis eu. Aliquam eleifend adipiscing leo sed cursus. Vivamus ut dolor sit amet nunc ornare tristique at in purus. Sed ullamcorper, est sit amet convallis fermentum, felis mauris malesuada sapien, ut lobortis ligula tortor ut nisi. Nulla metus augue, dapibus a auctor in, fringilla id erat. Aliquam erat volutpat. Duis nec molestie justo. Vestibulum vulputate odio sed risus dignissim mattis. In mollis justo eu orci tincidunt condimentum. Duis et sapien nec ipsum placerat pretium vitae quis elit. Pellentesque laoreet sem vitae nulla adipiscing eleifend. Aliquam nec massa sem, at laoreet ante. Morbi sapien turpis, lacinia non rutrum sed, ullamcorper a nisi.

Aenean pulvinar semper ligula, nec fermentum libero sagittis non. Duis rhoncus imperdiet turpis, in mattis nisi blandit nec. Vestibulum in turpis arcu, molestie ornare dolor. Integer pellentesque dapibus laoreet. Vestibulum id risus a diam imperdiet porta. Sed ornare dictum nibh non pellentesque. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Nulla facilisi. Sed facilisis sem lobortis diam sollicitudin gravida. Quisque congue, dui ut eleifend blandit, nibh risus mattis risus, quis tristique eros lacus non eros. Pellentesque convallis euismod tincidunt. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Fusce congue felis leo. Pellentesque dictum ante sem. Vestibulum auctor leo nibh, sit amet rutrum est. Donec eu leo at arcu dictum eleifend id a leo.

Cras tristique pulvinar varius. Aenean eleifend, augue a consequat feugiat, libero nisl volutpat lectus, imperdiet dignissim sem nisl a nisi. Aenean blandit, sapien ac fermentum rutrum, dolor leo semper turpis, sed posuere mi dolor et mauris. Curabitur diam purus, adipiscing eu pretium eget, dignissim molestie felis. Integer et neque lorem. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec felis mauris, vulputate et feugiat quis, sodales ac sem. Cras a eros orci. In congue pulvinar elit, ornare scelerisque enim rhoncus nec. Donec a elit metus, sit amet lobortis mi. Donec iaculis massa at erat consequat a pretium nisi auctor.

Aenean blandit, ipsum eu volutpat interdum, nisi sapien semper nunc, id venenatis quam libero eget lorem. Phasellus laoreet semper tortor vel porttitor. Curabitur in orci elit. Sed eu conva-

llis elit. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam facilisis dapibus eros, scelerisque gravida nunc tristique nec. Donec bibendum rutrum nibh consequat vulputate. Donec gravida nibh condimentum quam convallis in ullamcorper purus fringilla. Aenean pharetra tellus a mauris semper vitae malesuada leo mollis. Donec ante justo, dictum vitae lobortis vel, accumsan sit amet odio. Etiam semper, nulla in laoreet ultrices, ligula quam sodales felis, eget scelerisque nunc felis quis dui. Curabitur lorem urna, rhoncus auctor malesuada sed, volutpat id tellus. Morbi vitae odio velit, a iaculis tellus. Suspendisse vehicula nulla quis erat ullamcorper ullamcorper. Nulla lacinia mattis convallis. Fusce eu purus est. Proin urna massa, fermentum accumsan gravida at, posuere at enim.

Fusce eu lectus massa, quis molestie elit. Aliquam mollis laoreet suscipit. Phasellus nec augue ut orci venenatis porta eget in est. Integer luctus nulla ut felis pharetra eu vestibulum quam tempus. In eget tellus dolor, id gravida diam. Fusce blandit pulvinar congue. Donec fermentum ligula et ipsum egestas ac vulputate eros hendrerit. Ut ac elit sem, at elementum odio. Nulla a ipsum nibh, eu posuere elit. Nullam tincidunt volutpat purus et pellentesque. Nulla ultrices rutrum neque id rhoncus. Mauris fringilla, turpis in rutrum cursus, mi ante tempor mi, eu condimentum turpis nisi quis arcu. Fusce arcu sapien, elementum quis porttitor in, pulvinar vitae turpis.

Aenean sem dolor, aliquet vel aliquet et, congue in lectus. Nunc accumsan posuere posuere. Aenean ut dignissim ante. Suspendisse vitae metus eget tellus auctor luctus. Quisque sed sapien nisi. Sed purus mauris, ornare non fringilla et, hendrerit nec orci. In imperdiet commodo orci, ac imperdiet nibh tempor nec. Mauris condimentum lacus vitae erat malesuada quis varius mi vulputate. Cras eu quam id dui auctor mattis. Cras bibendum, odio a mattis venenatis, ipsum erat tincidunt massa, quis pretium odio arcu quis dolor. Cras sagittis pulvinar elit eu posuere. Mauris faucibus mauris sed nulla luctus et tristique turpis iaculis. Vivamus euismod arcu nec est lacinia quis faucibus justo pharetra. Suspendisse ac ligula nulla. Curabitur risus lacus, malesuada sed imperdiet tristique, viverra eu quam.

Cras ac condimentum justo. Duis sollicitudin orci id nisi rhoncus egestas. Vestibulum sit amet elementum massa. Donec ac erat pretium velit tempor lacinia. Sed mi erat, sodales non volutpat ac, faucibus ut nulla. Phasellus nec commodo sem. Ut sollicitudin odio sit amet est commodo ut placerat risus facilisis. Sed egestas semper tellus non porta. Ut ullamcorper, justo in cursus sollicitudin, enim tortor mattis lorem, nec hendrerit orci odio sed massa. Aliquam consectetur auctor ante, varius commodo metus ullamcorper eu. Praesent et mi eros.

Nunc risus nisi, adipiscing at pretium sed, pharetra quis orci. Quisque purus mi, sodales id iaculis nec, suscipit sit amet neque. Etiam facilisis condimentum euismod. Nunc eros justo, vulputate a blandit a, imperdiet id eros. Nulla ornare magna pretium urna semper accumsan. Vivamus porttitor tempor felis eu fermentum. Proin adipiscing, ipsum vel feugiat rutrum, nunc mauris ornare augue, non tristique massa nisi eget dolor. Nullam lorem tortor, faucibus vel feugiat ut, facilisis sed velit. Integer tempor metus eu dolor eleifend sit amet feugiat arcu euismod. Sed nec vulputate augue. In ullamcorper vulputate odio non dignissim. Quisque nec libero elit. In cursus sodales nisi, ut porttitor sem consequat at. Sed et neque nulla, a dapibus elit. Cras fringilla orci quis purus pharetra sagittis. Morbi laoreet iaculis commodo.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Sed sed ante arcu, vel tempus urna. Phasellus ut sollicitudin dui. Sed fermentum congue lorem, et venenatis tellus tincidunt vitae. Nullam nec leo eu nisi blandit malesuada in eget ante. Pellentesque congue mollis congue. Suspendisse potenti. Donec at luctus orci. Aenean in tortor interdum tortor mollis rutrum. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae;

Nunc feugiat porttitor ipsum quis egestas. Duis tempor hendrerit vestibulum. Curabitur sagittis elit nec leo convallis elementum. Fusce vestibulum consequat metus nec venenatis. Pellentesque feugiat risus vitae enim facilisis quis adipiscing dui faucibus. Aenean interdum, nibh id placerat

pharetra, tellus dui eleifend felis, fringilla tincidunt enim nunc sed enim. Cras ullamcorper enim eu sem molestie eget vestibulum diam eleifend. Duis viverra, sapien eu molestie accumsan, mauris lorem laoreet odio, et placerat lacus ante luctus sapien. Integer placerat pulvinar purus, non imperdiet risus gravida condimentum. Nunc dolor nisi, hendrerit at feugiat sed, tempus a lorem. Vestibulum gravida, quam vel luctus malesuada, ipsum nisl tristique neque, id rhoncus nulla arcu in massa. Etiam elementum urna quis arcu auctor eu placerat dolor vulputate. Phasellus semper facilisis tellus, in rutrum leo dapibus quis. In hendrerit pulvinar mi nec mattis. Pellentesque ut tortor felis, vitae lacinia augue. Donec nec placerat lacus.

Nulla facilisis vehicula luctus. Duis ut nibh ac nunc ultrices eleifend id nec mauris. Mauris blandit augue nec quam pulvinar sit amet placerat tellus tincidunt. Mauris dapibus, felis quis rhoncus consequat, eros tellus commodo nibh, quis pulvinar nisl velit sed ante. Vestibulum quis tristique neque. Sed hendrerit euismod lorem, eu gravida sapien aliquam ut. Sed et ipsum sit amet leo imperdiet convallis id sit amet tortor. Nulla facilisi. Ut et libero id dui dictum molestie. Suspendisse tristique justo non dui faucibus tempor. Nulla ultrices enim neque. Phasellus eleifend hendrerit sem, sit amet feugiat elit ultrices eget. Proin ut est ligula, at pharetra tellus. Donec mollis, est et mattis tincidunt, massa nibh fermentum elit, quis fringilla mi justo sed lorem. Cras et ligula a nisi ultrices semper. Nulla id dolor felis, in dignissim massa. Nullam quis arcu a justo cursus iaculis. Suspendisse vestibulum, justo in sodales auctor, ipsum sem semper odio, quis imperdiet eros lorem ut elit. Ut auctor, arcu eu bibendum dignissim, nunc urna accumsan arcu, ullamcorper sagittis odio lectus id lorem.

Ut aliquet aliquet convallis. Nulla dignissim egestas aliquam. Nulla enim nisi, posuere vitae tincidunt eget, ultrices eu metus. Vestibulum faucibus congue tellus, eu congue lectus tristique quis. Sed a risus magna, id rutrum nulla. Nunc at felis vitae est pharetra mattis. In quam ipsum, semper in eleifend quis, sagittis eget leo. Fusce arcu ligula, mollis sed tempus sit amet, semper eget dolor. Donec nec mauris id tellus iaculis sodales. Ut pretium pharetra odio, vitae blandit augue porttitor eget. In purus justo, tincidunt ac consectetur vitae, feugiat vitae quam. In commodo odio consequat tortor lobortis sed luctus leo dapibus. Nulla consequat pulvinar tortor. Morbi euismod ipsum eget dui faucibus vel tempus nisl facilisis. Mauris aliquam dolor sit amet eros congue et mattis metus auctor. Proin id lacinia lectus. Praesent leo turpis, pulvinar sed lacinia in, fringilla quis augue.

Nulla commodo justo quis nulla gravida eget ornare velit aliquet. Ut semper pharetra dolor placerat ultricies. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Phasellus feugiat fermentum urna, ac convallis leo adipiscing nec. Sed augue augue, aliquet vitae ultrices nec, laoreet sit amet ante. Duis pharetra diam sed velit tempus et gravida felis auctor. Mauris facilisis ante in lacus bibendum eget malesuada erat gravida. Praesent rhoncus, risus quis dapibus pretium, sem elit blandit nisi, at consequat leo libero vitae nulla. Nunc eget convallis leo. Pellentesque ac eros et augue tempus commodo in nec justo. Curabitur non dolor non elit tincidunt aliquam. Phasellus elit sapien, interdum id molestie id, malesuada sed arcu. Aenean justo orci, aliquam vitae placerat vitae, ornare vitae augue. Nullam ultricies convallis magna in bibendum. Nulla sollicitudin risus eu lorem ultricies tincidunt. Nulla faucibus egestas quam ut malesuada. Donec vitae nisi neque, aliquam auctor lectus. Mauris nec ante ut turpis egestas fringilla. Etiam rutrum bibendum tellus at vehicula.

Donec varius iaculis mauris, vel iaculis arcu tempus at. Cras dapibus gravida tellus, nec tincidunt sapien hendrerit sit amet. Vestibulum euismod, quam et volutpat suscipit, nunc diam blandit lectus, nec pharetra lorem nunc et enim. Pellentesque ac massa diam, vel adipiscing nisi. Integer in dolor id urna accumsan euismod ut id tortor. Quisque aliquet lorem non lorem malesuada semper. In at diam mauris. Maecenas tincidunt sagittis mauris. Ut augue arcu, adipiscing ut pellentesque porttitor, fringilla at justo. Donec ipsum orci, pellentesque at commodo eu, mollis et purus. Nam varius ornare tincidunt. Suspendisse non erat auctor quam auctor gravida non nec

sem. Cras ut ante sit amet sem rutrum vestibulum. Fusce purus quam, semper ac imperdiet nec, accumsan sed odio. Quisque ligula est, pretium et volutpat eget, aliquam a turpis. Suspendisse potenti. Praesent molestie turpis a lorem consequat sollicitudin at eget tortor. Vivamus eu urna a mi vestibulum pellentesque quis malesuada sapien.
Nullam sapien tellus, iaculis in consequat a, commodo nec purus. In aliquam velit sit amet eros rutrum elementum. Maecenas in eros nec risus porta dignissim. Nulla sollicitudin, purus ut tempus mattis, elit nunc ultrices velit, et convallis turpis quam ac enim. Duis purus lorem, congue vel semper eget, commodo at nulla. Etiam a odio massa, eget sagittis nisl. Vestibulum ipsum tortor, auctor vel sagittis sed, consequat vitae ipsum. Ut varius tincidunt turpis eget dignissim. Vivamus ut felis dolor, vel sollicitudin arcu. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Capítol 3

Disseny del sistema

3.1 Mòduls

Capítol 4

Extracció dels continguts d'un PDF

La primera idea a l'hora d'abordar el nostre projecte va ser intentar extreure informació directament dels fitxers PDF dels quals es disposa.

4.1 Dificultats

Les principals dificultats són:

- Caràcters especials: com Unicode o lligadures
- Flux del text dins del fitxer

4.2 Programari existent

Tot hi haver-hi diverses utilitats que permeten l'extracció del contingut d'un fitxer PDF en forma de text pla o HTML, totes presenten problemes similars en els punts comentats a la secció anterior.
A l'apèndix A hi ha exemples de com queden els texts extrets de diferents documents PDF.

Capítol 5

Cerca de referències a Internet

5.1 Primera idea: *Google Scholar*

5.2 Resta de cercadors

Hem preparat el nostre cercador per tal d'utilitzar les APIs dels cercadors *Google*, *Yahoo* i *Bing* i hem
El principal avantatge és la
Un inconvenient, hi ha biblioteques virtuals que no estan indexades en aquests serveis.

5.3 Ajustaments

Podem ajustar la manera com es fan les cerques a partir de certs paràmetres que es detallen a continuació.
En moltes ocasions, el cercador *Bing* mostra resultats corresponents a *Microsoft Academic Search* (un projecte molt similar a *Google Scholar*). Aquestes pàgines, però, no mostren prou informació com per generar referències. Per tant, les hem d'ometre.

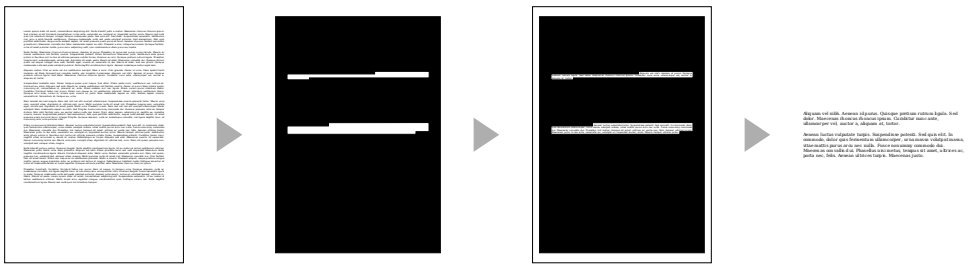
5.4 *Multithreading*

Un dels inconvenients que suposa el fet d'haver d'accedir a Internet, és la latència. Per reduir el temps que es perd esperant les dades, hem fet que l'aplicació creï diferents fils d'execució.

Capítol 6

Extracció d'Informació

En aquest capítol tractarem
En el nostre context, anomenarem *wrapper* a un troç de codi que podem utilitzar per extreure una peça d'informació concreta d'un document.
Podem imaginar-ho com filtre que només ens deixa veure una part del document que ens interessa.



Document Wrapper Apliquem el wrapper Text obtingut

6.1 *Wrappers* a mà

6.2 Inducció de *wrappers*

6.2.1 Generació automàtica de regles

Ruta d'un element HTML

Expressió regular

6.2.2 Avaluació dels *wrappers*

Una vegada hem generat el conjunt dels *wrappers* possibles per a un conjunt de documents, cal que avaluem quins d'ells funcionen millor. Utilitzem un sistema de vots positius i negatius i en calculem la mitjana amb la següent fórmula:

$$score = \frac{vots\ positius}{vots\ totals}$$

6.2.3 Reaprenentatge

El sistema està dissenyat per tal que, quan hi ha una davallada en el nombre de referències extretes correctament, provi de reaprendre els *wrappers* automàticament a partir dels exemples que té emmagatzemats d'execucions passades.

Capítol 7

Anàlisi de resultats

En aquest capítol es mostren les principals proves realitzades amb la nostra aplicació. Per cada prova s'explica el perquè dels resultats obtinguts.
A l'apèndix B es mostren tots els test que s'han dut a terme.

7.1 Només amb *wrappers* induïts

7.2 Utilitzant *wrappers* de referència

Capítol 8

Conclusions i Treball Futur

8.1 Objectius Assolits

8.2 Possibles Millores

Bibliografia

[Jr06] Nobody Jr. My article, 2006.

Apèndix A

Extracció Contingut PDF

Apèndix B

Resultats dels tests

A continuació es mostren els resultats complets de totes les proves realitzades a la nostra aplicació. L'explicació d'aquests números s'explica al capítol 7.