

Disseny i implementació d'una eina de traducció de documents \LaTeX a **HTML**

Albert Nadal i Garriga

Universitat de Lleida
Escola Universitària Politècnica
Enginyeria Tècnica en Informàtica de Gestió

Treball de Fi de Carrera

juliol de 2003

Fins on pot arribar el programari lliure? No hi ha límits, llevat quan lleis com els sistemes de patents prohibeixen el programari lliure en la seva totalitat. L'objectiu final és oferir programari lliure capaç de realitzar totes les tasques que els usuaris volen dur a terme, i per tant fer així el programari propietari obsolet...

Índex

1	Introducció	4
2	Conceptes preliminars	6
2.1	Què és el \LaTeX ?	6
2.2	Què és l'Hypertext?	8
2.3	Intruccions principals del llenguatge HTML	9
2.4	Traducció de \LaTeX a HTML: Per què?	11
2.5	Traducció de \LaTeX a HTML: Com?	12
2.6	Software i hardware utilitzat	13
3	Especificació d'una gramàtica per a la traducció	15
3.1	Analitzador lèxic Flex++	16
3.2	Analitzador sintàctic Bison++	17
3.3	Esquema d'interacció	18
3.4	Llistat de <i>tokens</i> i les macros que representen	21
4	Criteris, procediments i mètodes de traducció	31
4.1	Conjunt de macros reconegudes pel traductor	31
4.2	Estructura d'un document \LaTeX i HTML	32
4.3	Comptadors	33
4.4	Pila d'entorns	35
4.5	Accents i caràcters especials en mode text	36
4.6	Macros de funcionalitat específica	38
4.6.1	Text ordinari	38
4.6.2	Final de paràgraf	38
4.6.3	Trencament de línia	42
4.6.4	Línia horitzontal	43
4.6.5	Comentaris	43
4.7	Marques de format de text	43
4.7.1	Text emfatitzat	44
4.7.2	Text itàlica	44
4.7.3	Text mecanogràfic	44
4.7.4	Text negreta	44
4.7.5	Text subratllat	44

4.7.6	Text superscript	45
4.7.7	Text normal	45
4.7.8	Anidació de formats de text iguals	46
4.8	Subconjunt <code>maketitle</code> , <code>title</code> , <code>author</code> , <code>date</code> , <code>thanks</code> i <code>and</code> .	46
4.9	Notes a peu de pàgina. Macro <code>footnote</code>	48
4.10	Referències creuades. Subconjunt <code>label</code> i <code>ref</code>	51
4.11	Subconjunt <code>part</code> , <code>chapter</code> , <code>section</code> , <code>subsection</code> i <code>subsubsection</code> .	55
4.12	Subconjunt <code>paragraph</code> i <code>subparagraph</code>	60
4.13	Entorn <code>itemize</code> i <code>enumerate</code>	62
4.14	Entorn <code>tabbing</code>	63
4.15	Entorn <code>figure</code> i <code>table</code>	68
4.16	Macros <code>tableofcontents</code> , <code>listoffigures</code> i <code>listoftables</code> . . .	71
5	Conversió gràfica de fragments matemàtics i figures	75
5.1	Conversió de fragments matemàtics	75
5.2	Conversió de figures	77
6	Presentació i navegació	79
6.1	Entorn amb marcs	79
6.2	Sistema de navegació	82
7	Proves fetes	85
A	Manual d'aplicació per a l'usuari	105
A.1	Requeriments	105
A.2	Instal·lació	106
A.3	Utilització	106
A.4	Utilització del fitxer de configuració	106
A.5	Detecció d'errors	110
A.5.1	Errors d'estructura	110
A.5.2	Macros no reconegudes	110
A.5.3	Avisos	111
B	Conclusió i propostes per al futur	112
C	Preguntes més freqüents	114

Capítol 1

Introducció

A diferència dels programes de composició visual, en els quals l'escriptura, la composició, la visualització i la impressió d'un document es troben integrats en un únic programa informàtic, en els sistemes de tractament de textos científics aquestes tasques són realitzades per diferents components. El \LaTeX és en aquest sentit, el component que s'encarrega de la composició tipogràfica dels documents, mentre que un programa editor s'ocupa de l'escriptura del document, un traductor gràfic de pantalla permet la visualització o sortida per pantalla, i un traductor gràfic d'impressora permet la impressió o sortida impresa del document.

Així doncs, l'avantatge del sistema \LaTeX és clar. Sempre que es disposi dels estils necessaris o de persones que puguin programar aquests estils, els autors poden dedicar-se a escriure el contingut dels seus treballs científics sense haver de preocupar-se ni d'ocupar-se de qüestions sobre la forma de presentació dels treballs. És important doncs, saber diferenciar les necessitats de l'autor i la manera de satisfer-les. Es tracta d'apropar les eines informàtiques als autors, per evitar que aquests hagin de fer la tasca de tipògrafs.

En el cas de publicacions electròniques, no hi ha convencions editorials ben establertes, excepte en alguns casos en què les editorials han desenvolupat un conjunt de convencions per a la publicació electrònica, mitjançant els anomenats *llenguatges de marcatge estàndar generalitzats* (SGML). El cas més difós actualment és el de la publicació dels treballs científics a les xarxes globals d'ordinadors, mitjançant els *llenguatges de marcatge d'hipertext* (HTML).

En aquest projecte de final de carrera es proposa una eina de traducció que permeti, als autors de documents científics, la possibilitat d'adaptar els seus escrits realitzats en \LaTeX , per tal que aquests puguin formar part de la World Wide Web. En concret, aquesta eina és especialment útil per a aquelles persones familiaritzades amb el sistema \LaTeX , així com matemàtics, físics, informàtics, estudiants i en general a totes aquelles persones vinculades, d'alguna manera

o altra, en el món universitari i que tenen interès per compondre documents de qualitat.

A mode d'introducció es pot dir que la idea general del traductor consisteix en realitzar una rèplica fidel de l'estructura bàsica d'un document \LaTeX mitjançant la creació d'un conjunt de fitxers \HTML interconnectats entre si. Aquests fitxers poden ser explorats utilitzant els panells de navegació generats automàticament pel propi traductor. Així doncs, referències creuades, citacions, notes a peu de pàgina, taules de continguts, llistes de figures i llistes de taules són traduïts per mitjà d'enllaços hipertextuals, i tot el conjunt de macros \LaTeX reconegudes pel traductor són traduïdes a \HTML seguint uns rigorosos criteris de traducció amb l'objectiu de generar documents interactius que respectin fidelment el format original.

En aquesta memòria s'expliquen les principals característiques del traductor, així com el seu disseny, implementació, instal·lació i utilització entre d'altres. Primerament es defineixen uns conceptes preliminars que considero que són importants per a aquelles persones que encara no s'han familiaritzat amb el sistema \LaTeX o aquells que desconeixen la mecànica dels documents hipertextuals \HTML i pretenguin, a la llarga, fer ús del traductor. Tot i que la utilització del traductor és força senzilla i intuïtiva, recomano fer una lectura prèvia a aquesta memòria que l'acompanya. Amb això vull pretendre que l'usuari pugui treure'n el màxim rendiment del traductor i complir les seves necessitats amb un resultat satisfactori.

M'agradaria concloure aquesta introducció fent esment al tema de les actualitzacions, millores, notificació d'errades i altres aspectes futurs relacionats amb la documentació i implementació d'aquest traductor. El conjunt d'implementacions, documentacions, recull d'errades i notificacions d'errades que aniran sorgint en el transcurs del temps es pot trobar a l'apartat **GloboTeX - LaTeX to HTML translator**, de la web de l'autor, <http://arcana.ods.org>. Aquesta memòria (traduïda a \HTML amb el traductor proposat), també es pot consultar *online* a través de la web mencionada.

Agraïments La secció dels agraïments s'ha convertit ja en un estàndard fix en qualsevol treball final de carrera, i com que a aquestes alçades queda lleig fer un projecte sense agraïments aleshores faré un esforç i buscaré algú a qui donar les meves gràcies...

Dono les gràcies a la meva nòvia per corregir les faltes ortogràfiques d'aquesta memòria...

Uff! i també dono les gràcies a la família i als meus amics (<http://trugacanyes.cjb.net>).

Capítol 2

Conceptes preliminars

2.1 Què és el L^AT_EX?

T_EX és un sistema de composició de textos d'alta qualitat creat per Donald E. Knuth¹ al 1977, dirigit en particular a aquells textos que contenen una gran quantitat d'expressions matemàtiques. L'objectiu del sistema T_EX és la composició i impressió de documents que contenen text i fórmules matemàtiques, amb una qualitat similar a la que s'obté en un impremta. No s'ha de pensar en T_EX com un processador de paraules de tipus *WYSIWYG*².

La versió de T_EX que s'utilitza actualment³ és pràcticament igual a la de 1982, excepte per la inclusió d'un millor suport per a caràcters de 8 bits i múltiples idiomes al 1989.

T_EX es pronuncia [*tej*] i en un entorn ASCII s'escriu **Tex**

L^AT_EX és un paquet de macros per a T_EX, originalment escrit per Leslie Lamport per oferir un sistema de processament de documents científics. L^AT_EX proporciona una sèrie d'ordres per descriure l'estructura del document, de forma que l'usuari final no hagi de pensar gaire en la presentació. Mitjançant diferents classes de documents i paquets addicionals, el mateix document pot produir-se amb diferents aspectes.

La versió actual és *L^AT_EX2_ε*(1994) i es pronuncia [*latej*].

A diferència dels programes de composició visual, en els quals l'escriptura, la composició, la visualització i la impressió d'un document es troben integrades en un únic programa informàtic, en els sistemes de tractament de textos científics

¹<http://www-cs-staff.stanford.edu/~knuth/>

² *What You See Is What You Get*

³(3,14159)

aquestes tasques són realitzades per diferents components. \LaTeX és, en aquest sentit, el component que s'encarrega de la composició tipogràfica dels documents, mentre que un programa *editor* s'ocupa de l'escriptura del document, un traductor gràfic de pantalla permet la visualització o sortida per pantalla, i un traductor gràfic d'impressora permet la impressió o sortida impresa del document.

Un document \LaTeX és, en principi, un fitxer de text, s a dir, un fitxer que solament conté caràcters ASCII. Això vol dir que pràcticament qualsevol programa *editor* o processador de textos serveix per escriure un document \LaTeX . Aquests fitxers són els originals electrònics que serveixen d'entrada a un procés de composició, i normalment tenen l'extensió *.tex*.

Com a resultat del procés de composició o *compilació* mitjançant el sistema \LaTeX , s'obté un fitxer amb el mateix nom que el fitxer font, però amb extensió *.dvi*, el qual es pot visualitzar o imprimir amb l'ajuda de traductors gràfics apropiats.

Creació d'un document La creació d'un document consta dels següents passos.

1. Creació i modificació del fitxer font *.tex*. Es pot utilitzar qualsevol editor per crear o modificar aquest fitxer.
2. Compilació o composició. Es realitza mitjançant la comanda

latex fitxer.tex

Si hi ha errors s'ha de tornar al pas anterior. En cas contrari s'obté com a resultat el fitxer *.dvi*. És necessari executar dos vegades el *fitxer.tex* cada vegada que s'inclou una nova referència creuada o s'hagin efectuat canvis a la taula de continguts del document.

3. Visionat i/o impressió. El fitxer *.dvi* pot ser visualitzat per pantalla o imprès. Per a això s'utilitzen aplicacions com *xdvi*.
4. Acostuma a ser habitual convertir el fitxer *.dvi* a format *.ps*⁴ mitjançant la utilitat *dvips*.

dvips fitxer.dvi -o fitxer.ps

El fitxer obtingut pot ser enviat directament a una impressora **PostScript** o visualitzat en pantalla mitjançant algun programa com *gsview*⁵.

A les següents URL's hi ha disponibles extenses documentacions, així com manuals per a la composició de documents i distribucions gratuïtes de \LaTeX .

⁴PostScript

⁵<http://www.cs.wisc.edu/~ghost/>

- <http://www.math.hkbu.edu.hk/TeX/>
- <http://tex.loria.fr/english/general.html>
- <http://www.tug.org>
- <http://www.ctan.org>
- <http://tex.loria.fr/english/distributions.html>
- <ftp://ftp.dante.de/tex-archive>
- <ftp://ftp.tex.ac.uk/tex-archive>

2.2 Què és l'Hypertext?

La gran expansió que ha sofert Internet en aquests últims anys podem dir que és deguda a l'aparició de les "Planes Web". Actualment aquest és el servei més popular i utilitzat d'Internet. Consisteix en la publicació de documents que poden estar formats per text, gràfics, música, sons, animacions i altres recursos multimèdia. Aquests són del tipus hipermèdia, això vol dir que hi ha determinades paraules o imatges que són enllaços cap a altres documents de la xarxa.

Les paraules World Wide Web (La teranyina mundial), també conegudes com WWW, W3 o WEB, fan referència al conjunt abstracte d'informació que podem trobar al cyberespai, espai virtual format per Internet. El creador de la World Wide Web és Tim Bernes-Lee, un físic de la CERN, que en 1989 va començar a desenvolupar una forma senzilla de presentar informació multimèdia a través de la Xarxa.

El llenguatge utilitzat per crear les planes d'informació que podem veure quan naveguem per la WWW és el llenguatge HTML (HyperText Markup Language), un llenguatge de definició de planes derivat del SGML (Standard Generic Markup Language). Com tot llenguatge informàtic l'HTML ha sofert diverses revisions i evolucions. Ens trobem, doncs, davant d'un estàndard que està en procés de creixement accelerat, i això obliga a actualitzar constantment els navegadors que fem servir per visualitzar aquests documents. Actualment, el comitè que s'encarrega de regular totes les normes del llenguatge HTML és el World Wide Web Consortium, més conegut per les sigles W3C⁶.

Evolució del llenguatge HTML:

- La primera versió d'HTML, la 1.0 va ser desenvolupada en el CERN l'any 1992.
- L'HTML 2.0, desenvolupat per l'IETF's HTML working group, es va destacar per la inclusió del suport d'imatges i formularis, possibilitats no incloses en la versió 1.0 que només incloïa el suport per a enllaços.

⁶<http://www.w3.org/MarkUp/>

- L'HTML 3.0, que es basava en una proposta coneguda com HTML+ i que va començar a utilitzar-se de forma experimental l'any 1995, encara que van suposar una gran evolució mai no va ser acceptada com a estàndard. Va ser una millora d'aquest, l'HTML 3.2, el que finalment va rebre aquest reconeixement. El gran canvi que va suposar l'HTML 3.2 va ser la inclusió de gran quantitat de millores com les taules, applets, superscripts i subscripts i el que és més important, va permetre la inclusió de textos al voltant de les imatges. (cosa que havia suposat un gran problema fins aquell moment).
- L'HTML 4.0 és l'última versió d'HTML. Aquesta conté suport per a fulles d'estil, internacionalització, accessibilitat a la Web per a persones discapacitades i millores en els les taules, frames...
- L'HTML dinàmic, també conegut per les sigles DHTML (Dinamic HTML) suposa una revolució sobre la manera de crear les pàgines WEB. Està basat en una idea d'allò més simple: convertir les etiquetes tradicionals d'HTML en objectes programables, cosa que ens permet després poder manipular-les al nostre gust mitjançant JavaScript o altres llenguatges.

2.3 Intruccions principals del llenguatge HTML

Estructura de la plana i elements meta-informatius Abans de tot, hem d'analitzar els elements que componen l'estructura d'una plana Web. Una plana ha de començar sempre amb la instrucció `<HTML>` i finalitzarà amb `</HTML>`. Dins de la plana trobem dos elements bàsics, la capçalera i el cos de la plana. L'inici i el final de la capçalera es defineixen amb la instrucció `<HEAD>` i `</HEAD>` respectivament. La capçalera s'utilitza per a definir informació general sobre la plana. Dins la capçalera podem utilitzar els següents elements: `<TITLE>` defineix el títol de la plana, `<ISINDEX>` indicador per a la cerca, `<BASE>` adreça-base per a resoldre les URL relatives, `<META>` meta-informació addicional i, `<LINK>` que indica la relació existent entre dos planes o documents. El cos de la plana es defineix amb la instrucció `<BODY>` i `</BODY>`. Dins del cos de la plana situarem tots els elements que utilitzarem per definir el nostre document, text, imatges, enllaços, etc.

Inserció de paràgrafs i salts de línia `<Hn>` i `<P>` s'utilitzen per a mostrar text. `<Hn>` representa un títol (on n pot ser un número del 1 al 6) i `<P>` és un paràgraf. Les etiquetes `<HR>` i `
` permeten separar unitats lògiques d'un text, ja sigui amb una línia horitzontal o bé indicant un salt de línia. La instrucció `<BLOCKQUOTE>` s'utilitza per a definir un paràgraf que apareix sagnat respecte a la resta del text.

Inclusió de llistes Existeixen dos tipus de llistes: numerades i sense numerar. Les primeres tenen un número d'ordre davant de cada element (``, ``)

i les llistes sense numerar disposen d'una icona per a tal tasca. (,). Existeix un tercer format de llistes: les llistes de definicions. Les llistes de definicions estan formades per dos elements, el terme a definir i la seva definició. La instrucció <DL>serveix per crear la llista de definicions, <DT>representa el terme a definir i <DD>la definició.

Els enllaços Per a definir un enllaç (conector hipertextual) utilitzarem la comanda <A>, indicant a continuació amb la clàusula HREF la direcció a la qual fa referència l'enllaç (URL associada). També és possible crear un enllaç local si HREF apunta a un marcador, realitzat prèviament mitjançant la clàusula NAME.

Format i tipografia La clàusula ALING s'utilitza per a especificar la alineació del paràgraf i pot prendre els següents valors: center, left, right i (justify). De la mateixa manera que en un processador de textos, també podem escriure en negreta o en cursiva <I>. Una altra forma de resaltar el text és utilitzant el parapadeig <BLINK>. Finalment, per definir el tamany i el color del text emprem la instrucció seguida de les clàusules SIZE (pel tamany) i FACE (pel color).

Taules Les taules poden utilitzar-se per representar graelles de dades, com si es tractés d'un full de càlcul o per simular columnes. La instrucció <TABLE>s'utilitza per definir les propietats generals de la taula, com ara l'ample de la línia BORDER. Una vegada definida la taula, passarem a definir les files <TR>i per últim les cel·les que componen cada fila <TD>. L'atribut COLS especifica el número total de columnes de la taula.

Imatges La instrucció emprada per insertar una imatge en una plana Web és . La imatge s'especifica mitjançant la clàusula SRC tot assignant-li la direcció URL de la mateixa. Si la imatge es troba en el mateix directori o carpeta que la plana podem posar simplement el nom de la imatge. L'atribut ISMAP s'utilitza per definir enllaços o mapes dins d'una imatge.

Equacions Les següents intruccions permeten incloure equacions matemàtiques en la nostra plana (només es troben implemetades en HTML 3.2 o superior): <SUB>i <SUP>permeten definir un text que es tractarà com un subíndex o un superíndex respectivament, <MATH>engloba el conjunt d'instruccions i text que compona la fórmula, <BOX>conté una expressió formada per un numerador i un denominador, <OVER>separa el numerador del denominador, <ABOVE>conté un símbol que deu aparèixer sobre la expressió anterior i, <BELOW>conté un símbol que deu aparèixer sota la expressió anterior.

Formularis Un formulari és una plana Web en la qual l'usuari pot introduir informació que, posteriorment, rebrem per a que sigui processada per alguna aplicació o programa que es trobi al servidor. L'objecte <FORM>s'empra per

definir els límits del formulari dins d'una plana web. Les dos clàusules més importants d'aquesta instrucció són ACTION i METHOD, ja que determinen l'acció que es realitzarà i com es transmetran les dades. METHOD determina el format de dades del formulari i, ACTION és una cadena de caràcters amb la direcció URL on s'enviarà la informació del formulari. La instrucció <INPUT>ens permet crear caixes de text d'una sola línia. Per crear una llista, necessitarem dos instruccions: <SELECT>s'utilitza per definir la llista i <OPTION>per especificar els elements que la compondran.

Altres Per construir una plana amb marcs (els marcs permeten dividir la pantalla en diverses seccions i mostrar en cada una d'elles una plana web diferent) hem d'utilitzar les següents instruccions: <FRAMESET>, <FRAME> i <NOFRAME>. <FRAMESET>divideix la pantalla en columnes COLS o files ROWS, una vegada definides les files o columnes definim el contingut de cadascuna d'elles mitjançant l'element <FRAME>. La clàusula SRC indica la plana web que ha de carregar-se en el marc.

2.4 Traducció de L^AT_EX a HTML: Per què?

Són molts els motius pels quals pot ser necessari traduir documents en format L^AT_EX a HTML. Per una banda, els documents HTML disposen d'interactivitat, cosa que per a un document escrit en L^AT_EX aquest terme no existeix, però per altra banda, els documents L^AT_EX tenen una estructuració lògica i una elevada complexitat tipogràfica que en HTML és impensable.

Es tracta de combinar adequadament les possibilitats que ofereixen L^AT_EX i HTML per tal d'obtenir un document final interactiu que respecti fidelment el document original.

La traducció d'un document escrit en L^AT_EX a HTML suposa un increment considerable de les possibilitats que ofereixen les referències creuades, citacions, notes a peu de pàgina, taules de continguts, llistes de figures i llistes de taules, ja que aquests són traduïts a links hipertextuals. Així doncs, per donar una idea general, es pot dir que una simple referència creuada estàtica en L^AT_EX es converteix en una referència creuada dinàmica en HTML.

El traductor proposat es pot emprar com a...

- Eina de creació de llibres electrònics.
- Adaptació de documents científics per a la seva publicació internet.
- Actualització ràpida i còmoda de documents *on line*
- Eina per dotar d'un alt grau d'interactivitat documents seqüencials.

- Fàcil creació de *webs* d'estructuració complexa.

2.5 Traducció de \LaTeX a HTML: Com?

HTML i \LaTeX són llenguatges genèrics basats en etiquetes o marques (*tags*). Es pot dir que ambdós llenguatges són sintàcticament molt similars, ja que els dos han estat dissenyats amb la finalitat de crear documents.

Es tracta doncs, d'analitzar l'estructura sintàctica del document \LaTeX i traduir-la a la seva corresponent en HTML. La següent taula comparativa mostra les marques que tenen la mateixa funcionalitat tant en \LaTeX com en HTML. En molts casos les diferències són trivials, però en altres la traducció esdevé molt més complexa.

HTML	\LaTeX
<H1>	$\backslash\text{part}$ $\backslash\text{chapter}$
<H2>	$\backslash\text{section}$
<H3>	$\backslash\text{subsection}$
<H4>	$\backslash\text{subsubsection}$
<P>	par
<DL>	$\backslash\text{begin}\{\text{enumerate}\}$
	$\backslash\text{begin}\{\text{itemize}\}$
	$\backslash\text{item}$
<DL>	$\backslash\text{begin}\{\text{description}\}$
<DT>	$\backslash\text{item}$
<DD>	text
text	$\backslash\text{emph}\{\text{text}\}$
<I>text</I>	$\backslash\text{textit}\{\text{text}\}$
text	$\backslash\text{textbf}\{\text{text}\}$
<TT>text</TT>	$\backslash\text{texttt}\{\text{text}\}$

Així doncs, tenim que el següent fragment en \LaTeX ...

$\backslash\text{section}\{\textit{Hello world}\}$

es tradueix a HTML de la forma...

<H2>Hello world</H2>

El traductor llegeix del fitxer *.tex* origen, definit per l'usuari, i crea un conjunt de pàgines HTML que contenen la informació traduïda de l'original. La traducció es realitza de forma seqüencial, es a dir, a mesura que es va llegint del fitxer es va traduint, fins a arribar al final del fitxer⁷. Les macros i entorns \LaTeX són traduïts a HTML per la seva corresponent etiqueta, tal i com s'ha

⁷El byte EOF implica el final de la traducció

explicat a l'inici d'aquest apartat.

Existeixen situacions en les quals la traducció d'un fragment del document no és possible en HTML. Un clar exemple són els fragments matemàtics, la traducció dels quals suposa una complexitat elevada i en la majoria de casos la traducció no és possible. Davant d'aquestes situacions el traductor utilitza el propi compilador \LaTeX per convertir el fragment en format gràfic.

El format dels fitxers gràfics dels documents \LaTeX és **PostScript**, aquest format no és reconegut pels navegadors *web* estàndars, així doncs el traductor s'ha d'assegurar de convertir adequadament les imatges **PostScript** incloses en el document d'origen a un format gràfic que sigui vàlid per a la seva correcta visualització als navegadors *web*.

Al llarg d'aquesta documentació s'explica detalladament com ha estat implementat el traductor proposat i els criteris establerts per a la traducció.

2.6 Software i hardware utilitzat

Per al desenvolupament del traductor s'ha utilitzat el següent software de lliure distribució:

GNU/Flex++, A fast scanner generator Flex++ és una aplicació per generar analitzadors lèxics capaços de reconèixer models lèxics basats en uns patrons definits per l'usuari. Els analitzadors lèxics generats per Flex++ són codi en llenguatge C++, això permet que aquests puguin ser utilitzats per l'usuari com a objectes(classes).

GNU/Bison++, The YACC-compatible Parser Generator Bison++ és una aplicació per generar analitzadors sintàctics capaços d'analitzar la sintaxi d'una gramàtica definida per l'usuari. De la mateixa manera que el Flex++, Bison++ genera analitzadors en codi C++, per permetre que aquests puguin ser utilitzats per l'usuari com a objectes(classes). El fluxe d'entrada dels analitzadors sintàctics generats per Bison++ és el fluxe de sortida dels analitzadors lèxics generats per Flex++.

S'ha de dir que Flex++ i Bison++ són dos programes totalment independents, la qual cosa vol dir que ambdós poden funcionar perfectament si s'acoplen amb altres sistemes de generació d'analitzadors lèxics i sintàctics diferents.

La decisió d'haver escollit Flex++ i Bison++ es deguda a que els dos han estat implementats per funcionar acoplament amb la màxima compatibilitat mútua, a més a més perquè els dos generen *classes* analitzadores en C++, i per tant són útils per a implementar el traductor en C++.

GNU/g++ g++ és el compilador de c++ amb el qual s'ha implementat el traductor. D'aquesta manera el traductor és portable en qualsevol plataforma en la qual sigui possible compilar software desenvolupat en c++

TEX/L^ATEX El desenvolupament i el conjunt de proves realitzades per al traductor s'han fet amb les distribucions de L^ATEX `vtex` i `latex web2C 7.3.1`. Qualsevol distribució de L^ATEX és vàlida sempre i quan compleixi la normativa imposada pel creador del sistema TEX Donald E. Knuth.

Sistema El sistema operatiu utilitzat ha estat GNU/Linux instal·lat en un equip amb processador Pentium 3 650MHz amb 64MBytes de memòria RAM.

Capítol 3

Especificació d'una gramàtica per a la traducció

Per a poder traduir un document adequadament, el traductor ha de ser capaç de conèixer la gramàtica del llenguatge d'especificació \LaTeX . L'anàlisi de la gramàtica consta d'una anàlisi lèxica i una anàlisi sintàctica del llenguatge. Per a l'anàlisi lèxic, el traductor ha de conèixer el conjunt de macros de les que consta el llenguatge, per a l'anàlisi sintàctic el traductor ha de conèixer el conjunt d'estructures sintàctiques.

L'especificació d'una gramàtica permet al traductor validar si una determinada cadena pertany al llenguatge generat per la gramàtica. Per tal d'acceptar una seqüència de caràcters com una cadena generada per la gramàtica cal fer el següent:

- Analitzar-la i convertir-la en una cadena de símbols terminals de la gramàtica mitjançant un anàlisi lèxic.
- Analitzar la cadena de símbols terminals de la gramàtica fins a acceptar-la com a cadena vàlida si supera l'anàlisi sintàctic.

Per tant, per tal de realitzar una anàlisi lèxica i sintàctica als documents \LaTeX és necessari implementar ambdós analitzadors. Per evitar realitzar aquesta tasca, existeixen aplicacions generadores de programes capaços d'efectuar anàlisi lèxiques i sintàctiques a un fluxe de dades donat.

En definitiva, una anàlisi lèxica llegeix dades d'un fluxe d'entrada i el divideix en unitats lèxiques anomenades *tokens*, per tal que puguin ser processades per un altre programa. Normalment, aquest altre programa sol ser un analitzador sintàctic que s'encarrega d'analitzar gramaticalment la composició de l'entrada i determinar si aquesta coincideix o no amb una gramàtica definida.

3.1 Analitzador lèxic Flex++

Flex++ és un generador d'analitzadors lèxics. Els analitzadors es generen en forma de *classes* per al llenguatge C++, la qual cosa vol dir que poden ser usats en aplicacions implementades en aquest llenguatge de programació. Els analitzadors lèxics generats contenen un seguit d'operacions i atributs encapsulats, dels quals es convenient destacar els següents:

- **FILE* yyin** apunta a l'arxiu del qual Flex++ llegeix.
- **char* yytext** apunta a una cadena de caràcters que conté el valor semàntic del *token* en curs.
- **int yyleng** conté la longitud de la cadena apuntada per *yytext*.
- **int yylex()** Sempre que aquesta operació és executada, l'analitzador inicia la cerca del següent *token* del fitxer. *yylex()* segueix cercant *tokens* sempre que no s'arribi a la marca de final de fitxer (EOF) o be s'executi una operació *return*. En aquest últim cas, l'analitzador retorna el valor numèric del *token* que acaba de trobar i es manté inactiu fins que es torni a executar *yylex()*.

L'analitzador generat sotmet el fluxe de dades que va llegint del fitxer a un seguit de comparacions amb uns patrons definits per l'usuari. Quan una cadena compleix un patró, l'analitzador executa el fragment de codi C++ associat a aquell patró, el qual pot retornar, o no, un valor numèric que identifica el tipus de patró, aquest identificador s'anomena *token*. Tal i com s'ha explicat anteriorment, l'execució d'una instrucció *return* pausa el funcionament de l'analitzador, el qual tornarà a posar-se en marxa quan es cridi novament l'operació *yylex()*.

Cada vegada que una cadena de caràcters procedent del fluxe de dades d'entrada coincideix amb un patró, el valor semàntic de la cadena es guarda a memòria i *yytext* apunta a aquest fragment de memòria. El conjunt de patrons, declaracions i segments de codi C++ que l'usuari especifica per a l'analitzador lèxic es defineixen en un fitxer de text que ha de tenir la següent estructuració:

Declaracions i definicions

%%

Patrons i accions associades a cada patró

%%

Procediments auxiliars

La generació de l'analitzador lèxic s'efectua de la manera següent:

flex++ fitxer.l

Aquesta operació crea el fitxer *lex.yy.c*.

3.2 Analitzador sintàctic Bison++

Bison++ és un generador d'analitzador sintàctics. Els analitzadors es generen de la mateixa forma que amb el Flex++, és a dir, en forma de *classes* per al llenguatge C++. Els analitzadors sintàctics generats contenen un seguit d'operacions i atributs encapsulats, dels quals es convenient destacar els següents:

- **int yylex()** Els analitzadors sintàctics generats també disposen de la funció encapsulada *int yylex()*, tanmateix, aquesta operació està buida i és l'usuari qui ha d'implementar-la.
- **void yyparse()** Sempre que aquesta operació és executada, l'analitzador crida la funció encapsulada *int yylex()*. Amb la crida d'aquesta operació, l'analitzador obté els identificadors de *tokens* retornats per l'analitzador lèxic.
- **struct yyval** Aquesta estructura de dades ha de ser definida per l'usuari. *yyval* emmagatzema el valor semàntic del *token* rebut. *yyval* s'ha d'actualitzar cada vegada que s'executa la funció *int yylex()*.

Una especificació sintàctica per a Bison++ consta d'un fitxer segmentat en la següent estructuració:

%{ Operacions addicionals de l'usuari **%}**

%name Nom de la classe

%union{ Definició de l'estructura *struct yyval* **}**

%define MEMBERS Definició d'operacions membre de la classe

%header{ Definició de la capçalera de la classe i implementació d'operacions específiques **}**

%token <tipus de dada> Conjunt d'identificadors de *token* que es capaç de retornar l'analitzador lèxic. Els *tokens* inclosos aquí tenen un valor semàntic del tipus indicat a l'inici de la línia.

%token Conjunt d'identificadors de *token* que es capaç de retornar l'analitzador lèxic. En aquest cas els *tokens* mencionats no contenen valor semàntic.

%type <tipus de dada> Conjunt de símbols no terminals (estructures sintàctiques) que retornen un valor semàntic del tipus indicat a l'inici de la línia.

%start Símbol no terminal inicial amb el qual l'analitzador sintàctic començarà el procés.

%%

Aquí s'inclouen les especificacions i implementacions de les estructures sintàctiques de l'analitzador i les accions que realitzen.

%%

La generació de l'analitzador sintàctic s'efectua de la manera següent:

bison++ fitxer.y

Aquesta operació crea el fitxer *fitxer.tab.c*.

3.3 Esquema d'interacció

L'acoplament de l'analitzador lèxic i el sintàctic és fonamental per tal de crear la gramàtica adequadament. Existeixen diverses maneres d'acoplar ambdós analitzadors, de les quals he escollit el mètode que explico a continuació.

Tal i com s'ha explicat en l'apartat 3.2, els analitzadors sintàctics generats contenen l'operació encapsulada *void yyparse()* i la funció virtual encapsulada *int yylex()*.

yyparse() inicia l'anàlisi, i sempre que aquest necessita un *token* el demana fent una crida a la funció *int yylex()*. Tanmateix, *int yylex()* és una funció buida i no pot ser implementada directament per l'usuari, ja que les operacions encapsulades dels analitzador generats per Bison++ no es poden modificar. Aquest problema es resol implementant una *classe* adaptadora *subclasse* de l'analitzador sintàctic.

```
class GloboTex : public AnalitzadorSintactic
{
private:
    MyScanner lexic;
    :
    virtual int yylex()
    {
        int token = lexic.yylex();
        yylval.valor = (char*)malloc(strlen(lexic.yytext)+1);
        sprintf(yylval.valor,"%s",lexic.yytext);
        return token;
    };
    :
}
```

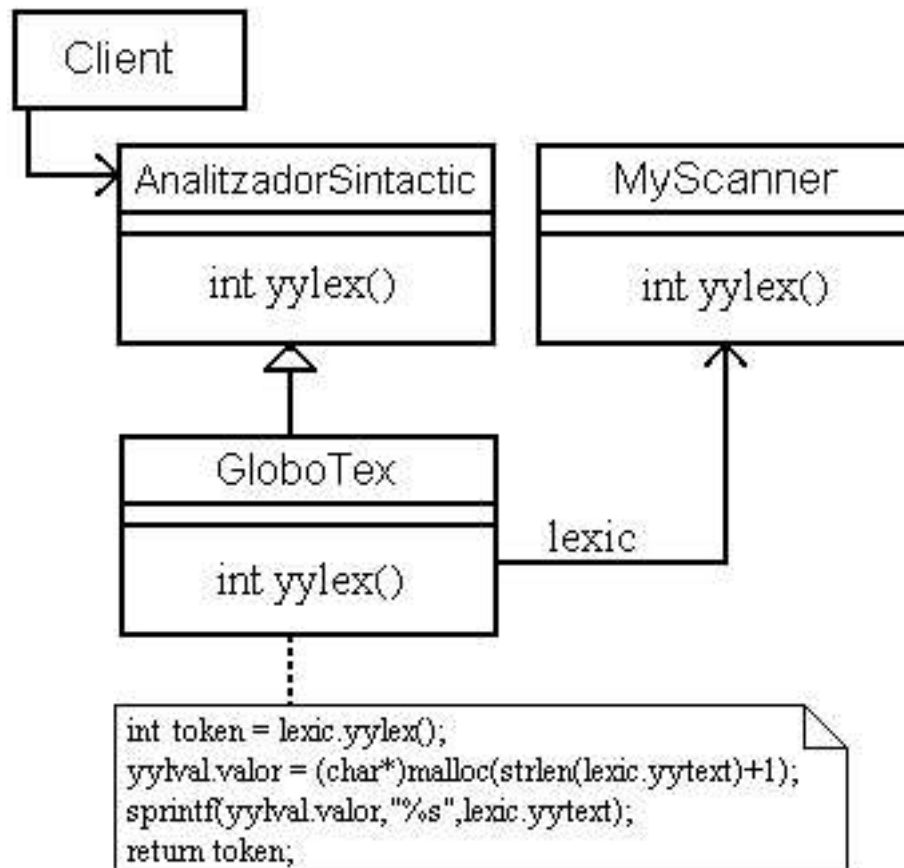


Figura 3.1: Esquema

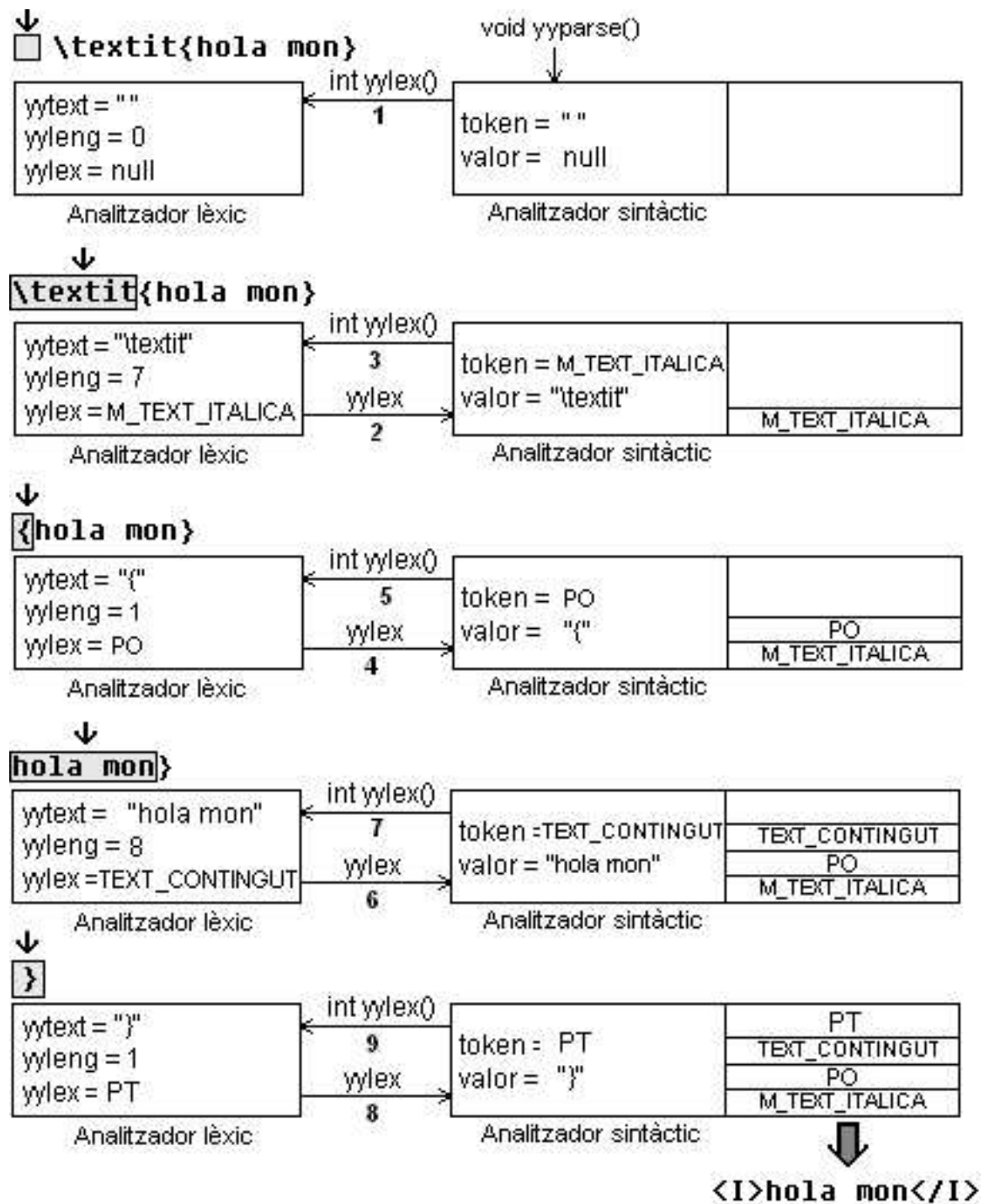


Figura 3.2: Exemple d'interacció durant el procés de traducció

A la figura 3.1 es pot veure com la *classe* adaptadora hereta la funció *intyylex()*, i aquesta, al tractar-se d'una funció virtual, pot ser aleshores implementada per enllaçar l'analitzador lèxic *MyScanner*.

3.4 Llistat de *tokens* i les macros que representen

Les següents taules mostren els conjunt de macros suportades pel traductor amb els seus respectius patrons lèxics i identificadors de *token* associats. Les macros que no tenen identificador de *token* associat són cadenes de caràcters que no tenen efecte a nivell sintàctic: comentaris i canvis de línia.

- Taula 3.1: Macros de funcionalitat específica
- Taula 3.2: Macros de format de text
- Taula 3.3: Macros de tamany de text
- Taula 3.4: Macros de funcionalitat estructural
- Taula 3.5: Macros de símbols i caràcters especials
- Taula 3.6: Macros d'entorn
- Taula 3.7: Macros d'accents i signes diacrítics sobre la lletra *a*
- Taula 3.8: Macros d'accents i signes diacrítics sobre la lletra *e*
- Taula 3.9: Macros d'accents i signes diacrítics sobre la lletra *i*
- Taula 3.10: Macros d'accents i signes diacrítics sobre la lletra *o*
- Taula 3.11: Macros d'accents i signes diacrítics sobre la lletra *u*
- Taula 3.12: Macros de lletres específiques

L^AT_EX	Patró	Id. Token
text ordinari	$[\wedge\\n\{\}\[\]\%$,]*$	TEXT_CONTINGUT
{	"{"	PO
}	"}"	PT
["["	CO
]	"]"	CT
\\	\\	FP
\par	"\par"	FP
\linebreak	"\linebreak"	NL
\newline	"\newline"	NL
\rule	"\rule"	LINIA
$\$text\$$	"\$"[$\wedge\n]"\$"	MATEMATIC
$\$\$text\$\$$	"\$"\$[$\wedge\$n$]"\$"\$	MATEMATIC
%comentari\n	"%"[$\wedge n$]* \n	
\n	"\n"	
\includegraphics	"\includegraphics"	IMATGE
\epsfig	"\epsfig"	IMATGEFIG
\=	"\="	POS_TAB
\>	"\>"	TAB
\hline	"\hline"	HLINE

Taula 3.1: Macros de funcionalitat específica

L^AT_EX	Patró	Id. Token
\emph	"\emph"	M_TEXT_EMF
\textnormal	"\textnormal"	M_TEXT_NORMAL
\textit	"\textit"	M_TEXT_ITALICA
\textbf	"\textbf"	M_TEXT_BOLD
\underline	"\underline"	M_TEXT_SUBRATLLAT
\textsuperscript	"\textsuperscript"	M_TEXT_SUP
\texttt	"\texttt"	M_TEXT_MECANOGRAFIC
\it	"\it"	D_TEXT_IT
\bf	"\bf"	D_TEXT_BF
\em	"\em"	D_TEXT_EM
\tt	"\tt"	D_TEXT_TT

Taula 3.2: Macros de format de text

L^AT_EX	Patró	Id. Token
<code>\tiny</code>	<code>"\\tiny"</code>	D_TEXT_TINY
<code>\scriptsize</code>	<code>"\\scriptsize"</code>	D_TEXT_SCRIPTSIZE
<code>\footnotesize</code>	<code>"\\footnotesize"</code>	D_TEXT_FOOTNOTESIZE
<code>\small</code>	<code>"\\small"</code>	D_TEXT_SMALL
<code>\normalsize</code>	<code>"\\normalsize"</code>	D_TEXT_NORMALSIZE
<code>\large</code>	<code>"\\large"</code>	D_TEXT_LARGE
<code>\Large</code>	<code>"\\Large"</code>	D_TEXT_LLARGE
<code>\LARGE</code>	<code>"\\LARGE"</code>	D_TEXT_LLLARGE
<code>\huge</code>	<code>"\\huge"</code>	D_TEXT_HUGE
<code>\Huge</code>	<code>"\\Huge"</code>	D_TEXT_HHUGE

Taula 3.3: Macros de tamany de text

L^AT_EX	Patró	Id. Token
<code>\documentclass</code>	<code>"\\documentclass"</code>	M_DOC_CLASS
<code>\usepackage</code>	<code>"\\usepackage"</code>	M_PACKAGE
<code>\section</code>	<code>"\\section"</code>	M_SECCIO
<code>\subsection</code>	<code>"\\subsection"</code>	M_SUBSECCIO
<code>\subsubsection</code>	<code>"\\subsubsection"</code>	M_SUBSUBSECCIO
<code>\chapter</code>	<code>"\\chapter"</code>	M_CAPITOL
<code>\part</code>	<code>"\\part"</code>	M_PART
<code>\item</code>	<code>"\\item"</code>	M_ITEM
<code>\subitem</code>	<code>"\\subitem"</code>	M_SUBITEM
<code>\footnote</code>	<code>"\\footnote"</code>	M_PEU_DE_PAGINA
<code>\tableofcontents</code>	<code>"\\tableofcontents"</code>	M_TAUOLA_DE_CONTINGUTS
<code>\listoffigures</code>	<code>"\\listoffigures"</code>	M_LLISTA_DE_FIGURES
<code>\listoftables</code>	<code>"\\listoftables"</code>	M_LLISTA_DE_TAULES
<code>\label</code>	<code>"\\label"</code>	M_ETIQUETA
<code>\ref</code>	<code>"\\ref"</code>	M_REFERENCIA
<code>\paragraph</code>	<code>"\\paragraph"</code>	M_PARAGRAF
<code>\subparagraph</code>	<code>"\\subparagraph"</code>	M_SUBPARAGRAF
<code>\caption</code>	<code>"\\caption"</code>	M_CAPTION
<code>\title</code>	<code>"\\title"</code>	M_TITOL
<code>\author</code>	<code>"\\author"</code>	M_AUTOR
<code>\and</code>	<code>"\\and"</code>	M_AND
<code>\date</code>	<code>"\\date"</code>	M_DATA
<code>\thanks</code>	<code>"\\thanks"</code>	M_THANKS
<code>\maketitle</code>	<code>"\\maketitle"</code>	M_FER_TITOL
<code>\cite</code>	<code>"\\cite"</code>	M_CITACIO
<code>\bibitem</code>	<code>"\\bibitem"</code>	M_ITEM_BIBLIOGRAFIC

Taula 3.4: Macros de funcionalitat estructural

L^AT_EX	Patró	Id. Token
$\backslash \$$	" $\backslash \$$ "	S_DOLLAR
$\backslash \text{textdollar}$	" $\backslash \text{textdollar}$ "	S_DOLLAR
$\backslash \&$	" $\backslash \&$ "	S_AMPERSAND
$\backslash \%$	" $\backslash \%$ "	S_TANTPERCENT
$\backslash -$	" $\backslash -$ "	S_GUIOINF
$\backslash \text{textunderscore}$	" $\backslash \text{textunderscore}$ "	S_GUIOINF
$\backslash \{$	" $\backslash \{$ "	S_CLAUOBERTA
$\backslash \text{textbraceleft}$	" $\backslash \text{textbraceleft}$ "	S_CLAUOBERTA
$\backslash \}$	" $\backslash \}$ "	S_CLAUTANCADA
$\backslash \text{textbraceright}$	" $\backslash \text{textbraceright}$ "	S_CLAUTANCADA
$\backslash \text{textless}$	" $\backslash \text{textless}$ "	S_MENORQUE
$\backslash \text{textgreater}$	" $\backslash \text{textgreater}$ "	S_MAJORQUE
$\backslash \text{neq}$	" $\backslash \text{neq}$ "	S_DIFERENT
$\backslash \text{leq}$	" $\backslash \text{leq}$ "	S_MENOR_IGUAL
$\backslash \text{geq}$	" $\backslash \text{geq}$ "	S_MAJOR_IGUAL
$\backslash \text{approx}$	" $\backslash \text{approx}$ "	S_APROXIMANT
$\backslash \text{equiv}$	" $\backslash \text{equiv}$ "	S_IDENTIC
$\backslash \text{infty}$	" $\backslash \text{infty}$ "	S_INFINIT
$\backslash \text{cap}$	" $\backslash \text{cap}$ "	S_INTERSECCIO
$\backslash \text{int}$	" $\backslash \text{int}$ "	S_INTEGRAL
$\backslash \text{surd}$	" $\backslash \text{surd}$ "	S_ARREL_QUADRADA
$\backslash \text{sum}$	" $\backslash \text{sum}$ "	S_SUMATORI
$\backslash \text{prod}$	" $\backslash \text{prod}$ "	S_PRODUCTORI
$\backslash \text{triangle}$	" $\backslash \text{triangle}$ "	S_INCREMENT
$\backslash \text{partial}$	" $\backslash \text{partial}$ "	S_DIFERENCIAL
$\backslash \text{textbackslash}$	" $\backslash \text{textbackslash}$ "	S_CONTRABARRA
$\backslash \text{textbar}$	" $\backslash \text{textbar}$ "	S_BARRA
$\backslash \text{textbullet}$	" $\backslash \text{textbullet}$ "	S_PUNT
$\backslash \text{textdaggerdbl}$	" $\backslash \text{textdaggerdbl}$ "	S_DOBLECREUETA
$\backslash \text{textdagger}$	" $\backslash \text{textdagger}$ "	S_CREUETA
$\backslash \text{textparagraph}$	" $\backslash \text{textparagraph}$ "	S_PARAGRAF
$\backslash \text{textsection}$	" $\backslash \text{textsection}$ "	S_SECCIO
$\backslash \text{textcopyright}$	" $\backslash \text{textcopyright}$ "	S_COPYRIGHT
$\backslash \text{textasciicircum}$	" $\backslash \text{textasciicircum}$ "	S_CIRCUM
$\backslash \text{textasciitilde}$	" $\backslash \text{textasciitilde}$ "	S_TILDE
$\backslash \text{textregistered}$	" $\backslash \text{textregistered}$ "	S_REGISTRAT
$\backslash \text{texttrademark}$	" $\backslash \text{texttrademark}$ "	S_TRADEMARK
$\backslash \text{textordfeminine}$	" $\backslash \text{textordfeminine}$ "	S_AFEM
$\backslash \text{textordmasculine}$	" $\backslash \text{textordmasculine}$ "	S_OMASC
$\backslash \text{ss}$	" $\backslash \text{ss}$ "	S_SHARP

Taula 3.5: Macros de símbols i caràcters especials

L^AT_EX	Patró	Id. Token
<code>\begin{document}</code>	<code>"\\begin{document}"</code>	INICI_DOC
<code>\end{document}</code>	<code>"\\end{document}"</code>	FI_DOC
<code>\begin{titlepage}</code>	<code>"\\begin{titlepage}"</code>	INICI_TITOL
<code>\end{titlepage}</code>	<code>"\\end{titlepage}"</code>	FI_TITOL
<code>\begin{enumerate}</code>	<code>"\\begin{enumerate}"</code>	INICI_ENUMERAT
<code>\end{enumerate}</code>	<code>"\\end{enumerate}"</code>	FI_ENUMERAT
<code>\begin{figure}</code>	<code>"\\begin{figure}"</code>	INICI_FIGURA
<code>\end{figure}</code>	<code>"\\end{figure}"</code>	FI_FIGURA
<code>\begin{table}</code>	<code>"\\begin{table}"</code>	INICI_TAUOLA
<code>\end{table}</code>	<code>"\\end{table}"</code>	FI_TAUOLA
<code>\begin{abstract}</code>	<code>"\\begin{abstract}"</code>	INICI_ABSTRACT
<code>\end{abstract}</code>	<code>"\\end{abstract}"</code>	FI_ABSTRACT
<code>\begin{tabular}</code>	<code>"\\begin{tabular}"</code>	INICI_TABULAR
<code>\end{tabular}</code>	<code>"\\end{tabular}"</code>	FI_TABULAR
<code>\begin{equation}</code>	<code>"\\begin{equation}"</code>	INICI_EQUACIO
<code>\end{equation}</code>	<code>"\\end{equation}"</code>	FI_EQUACIO
<code>\begin{itemize}</code>	<code>"\\begin{itemize}"</code>	INICI_ITEMIZE
<code>\end{itemize}</code>	<code>"\\end{itemize}"</code>	FI_ITEMIZE
<code>\begin{array}</code>	<code>"\\begin{array}"</code>	INICI_ARRAY
<code>\end{array}</code>	<code>"\\end{table}"</code>	FI_ARRAY
<code>\begin{center}</code>	<code>"\\begin{center}"</code>	INICI_CENTRAT
<code>\end{center}</code>	<code>"\\end{center}"</code>	FI_CENTRAT
<code>\begin{flushleft}</code>	<code>"\\begin{flushleft}"</code>	INICI_TEXT_ESQ
<code>\end{flushleft}</code>	<code>"\\end{flushleft}"</code>	FI_TEXT_ESQ
<code>\begin{flushright}</code>	<code>"\\begin{flushright}"</code>	INICI_TEXT_DRETA
<code>\end{flushright}</code>	<code>"\\end{flushright}"</code>	FI_TEXT_DRETA
<code>\begin{tabbing}</code>	<code>"\\begin{tabbing}"</code>	INICI_TABBING
<code>\end{tabbing}</code>	<code>"\\end{tabbing}"</code>	FI_TABBING
<code>\begin{thebibliography}</code>	<code>"\\begin{thebibliography}"</code>	INICI_BIBLIOGRAFIA
<code>\end{thebibliography}</code>	<code>"\\end{thebibliography}"</code>	FI_BIBLIOGRAFIA

Taula 3.6: Macros d'entorn

L^AT_EX	Patró	Id. Token
<code>\`{a}</code>	<code>"\\`{a}" ""\\grave{a}"</code>	<code>A_A_OBERT</code>
<code>\grave{a}</code>	<code>"\\`{a}" ""\\grave{a}"</code>	<code>A_A_OBERT</code>
<code>\' {a}</code>	<code>"\\\'{a}" ""\\acute{a}"</code>	<code>A_A_TANCAT</code>
<code>\acute{a}</code>	<code>"\\\'{a}" ""\\acute{a}"</code>	<code>A_A_TANCAT</code>
<code>\^{a}</code>	<code>"\\^{a}" ""\\hat{a}"</code>	<code>A_A_CIRCUM</code>
<code>\hat{a}</code>	<code>"\\^{a}" ""\\hat{a}"</code>	<code>A_A_CIRCUM</code>
<code>\" {a}</code>	<code>"\\\"{a}" ""\\ddot{a}"</code>	<code>A_A_DIERESIS</code>
<code>\ddot{a}</code>	<code>"\\\"{a}" ""\\ddot{a}"</code>	<code>A_A_DIERESIS</code>
<code>\ ~{a}</code>	<code>"\\ \~{a}" ""\\tilde{a}"</code>	<code>A_A_TILDE</code>
<code>\tilde{a}</code>	<code>"\\ \~{a}" ""\\tilde{a}"</code>	<code>A_A_TILDE</code>
<code>\={a}</code>	<code>"\\={a}" ""\\bar{a}"</code>	<code>A_A_MACRON</code>
<code>\bar{a}</code>	<code>"\\={a}" ""\\bar{a}"</code>	<code>A_A_MACRON</code>
<code>\u{a}</code>	<code>"\\u{a}" ""\\breve{a}"</code>	<code>A_A_BREVE</code>
<code>\breve{a}</code>	<code>"\\u{a}" ""\\breve{a}"</code>	<code>A_A_BREVE</code>
<code>\r{a}</code>	<code>"\\r{a}"</code>	<code>A_A_RING</code>
<code>\ae</code>	<code>"\\ae"</code>	<code>A_A_AE_DIFTONG</code>
<code>\`{A}</code>	<code>"\\`{A}" ""\\grave{A}"</code>	<code>A_A_OBERT_M</code>
<code>\grave{A}</code>	<code>"\\`{A}" ""\\grave{A}"</code>	<code>A_A_OBERT_M</code>
<code>\' {A}</code>	<code>"\\\'{A}" ""\\acute{A}"</code>	<code>A_A_TANCAT_M</code>
<code>\acute{A}</code>	<code>"\\\'{A}" ""\\acute{A}"</code>	<code>A_A_TANCAT_M</code>
<code>\^{A}</code>	<code>"\\^{A}" ""\\hat{A}"</code>	<code>A_A_CIRCUM_M</code>
<code>\hat{A}</code>	<code>"\\^{A}" ""\\hat{A}"</code>	<code>A_A_CIRCUM_M</code>
<code>\" {A}</code>	<code>"\\\"{A}" ""\\ddot{A}"</code>	<code>A_A_DIERESIS_M</code>
<code>\ddot{A}</code>	<code>"\\\"{A}" ""\\ddot{A}"</code>	<code>A_A_DIERESIS_M</code>
<code>\ ~{A}</code>	<code>"\\ \~{A}" ""\\tilde{A}"</code>	<code>A_A_TILDE_M</code>
<code>\tilde{A}</code>	<code>"\\ \~{A}" ""\\tilde{A}"</code>	<code>A_A_TILDE_M</code>
<code>\={A}</code>	<code>"\\={A}" ""\\bar{A}"</code>	<code>A_A_MACRON_M</code>
<code>\bar{A}</code>	<code>"\\={A}" ""\\bar{A}"</code>	<code>A_A_MACRON_M</code>
<code>\u{A}</code>	<code>"\\u{A}" ""\\breve{A}"</code>	<code>A_A_BREVE_M</code>
<code>\breve{A}</code>	<code>"\\u{A}" ""\\breve{A}"</code>	<code>A_A_BREVE_M</code>
<code>\r{A}</code>	<code>"\\r{A}"</code>	<code>A_A_RING_M</code>
<code>\AE</code>	<code>"\\AE"</code>	<code>A_A_AE_DIFTONG_M</code>

Taula 3.7: Macros d'accents i signes diacrítics sobre la lletra *a*

L^AT_EX	Patró	Id. Token
<code>\`{e}</code>	<code>"\\`{e}" "\\grave{e}"</code>	<code>A_E_OBERT</code>
<code>\grave{e}</code>	<code>"\\`{e}" "\\grave{e}"</code>	<code>A_E_OBERT</code>
<code>\`{e}</code>	<code>"\\\'{e}" "\\acute{e}"</code>	<code>A_E_TANCAT</code>
<code>\acute{e}</code>	<code>"\\\'{e}" "\\acute{e}"</code>	<code>A_E_TANCAT</code>
<code>\^{e}</code>	<code>"\\^ {e}" "\\hat{e}"</code>	<code>A_E_CIRCUM</code>
<code>\hat{e}</code>	<code>"\\^ {e}" "\\hat{e}"</code>	<code>A_E_CIRCUM</code>
<code>\`{e}</code>	<code>"\\\"{e}" "\\ddot{e}"</code>	<code>A_E_DIERESIS</code>
<code>\ddot{e}</code>	<code>"\\\"{e}" "\\ddot{e}"</code>	<code>A_E_DIERESIS</code>
<code>\={e}</code>	<code>"\\={e}" "\\bar{e}"</code>	<code>A_E_MACRON</code>
<code>\bar{e}</code>	<code>"\\={e}" "\\bar{e}"</code>	<code>A_E_MACRON</code>
<code>\.{e}</code>	<code>"\\\. {e}" "\\dot{e}"</code>	<code>A_E_DOT</code>
<code>\dot{e}</code>	<code>"\\\. {e}" "\\dot{e}"</code>	<code>A_E_DOT</code>
<code>\u{e}</code>	<code>"\\u {e}" "\\breve{e}"</code>	<code>A_E_BREVE</code>
<code>\breve{e}</code>	<code>"\\u {e}" "\\breve{e}"</code>	<code>A_E_BREVE</code>
<code>\v{e}</code>	<code>"\\v {e}" "\\check{e}"</code>	<code>A_E_CHECK</code>
<code>\check{e}</code>	<code>"\\v {e}" "\\check{e}"</code>	<code>A_E_CHECK</code>
<code>\`{E}</code>	<code>"\\`{E}" "\\grave{E}"</code>	<code>A_E_OBERT_M</code>
<code>\grave{E}</code>	<code>"\\`{E}" "\\grave{E}"</code>	<code>A_E_OBERT_M</code>
<code>\`{E}</code>	<code>"\\\'{E}" "\\acute{E}"</code>	<code>A_E_TANCAT_M</code>
<code>\acute{E}</code>	<code>"\\\'{E}" "\\acute{E}"</code>	<code>A_E_TANCAT_M</code>
<code>\^{E}</code>	<code>"\\^ {E}" "\\hat{E}"</code>	<code>A_E_CIRCUM_M</code>
<code>\hat{E}</code>	<code>"\\^ {E}" "\\hat{E}"</code>	<code>A_E_CIRCUM_M</code>
<code>\`{E}</code>	<code>"\\\"{E}" "\\ddot{E}"</code>	<code>A_E_DIERESIS_M</code>
<code>\ddot{E}</code>	<code>"\\\"{E}" "\\ddot{E}"</code>	<code>A_E_DIERESIS_M</code>
<code>\={E}</code>	<code>"\\={E}" "\\bar{E}"</code>	<code>A_E_MACRON_M</code>
<code>\bar{E}</code>	<code>"\\={E}" "\\bar{E}"</code>	<code>A_E_MACRON_M</code>
<code>\.{E}</code>	<code>"\\\. {E}" "\\dot{E}"</code>	<code>A_E_DOT_M</code>
<code>\dot{E}</code>	<code>"\\\. {E}" "\\dot{E}"</code>	<code>A_E_DOT_M</code>
<code>\u{E}</code>	<code>"\\u {E}" "\\breve{E}"</code>	<code>A_E_BREVE_M</code>
<code>\breve{E}</code>	<code>"\\u {E}" "\\breve{E}"</code>	<code>A_E_BREVE_M</code>
<code>\v{E}</code>	<code>"\\v {E}" "\\check{E}"</code>	<code>A_E_CHECK_M</code>
<code>\check{E}</code>	<code>"\\v {E}" "\\check{E}"</code>	<code>A_E_CHECK_M</code>

Taula 3.8: Macros d'accents i signes diacrítics sobre la lletra *e*

L^AT_EX	Patró	Id. Token
\'{\i}	"\\'\{\i}" \grave{\i}"	A_I_OBERT
\grave{\i}	"\\'\{\i}" \grave{\i}"	A_I_OBERT
\' {\i}	"\\'\{\i}" \acute{\i}"	A_I_TANCAT
\acute{\i}	"\\'\{\i}" \acute{\i}"	A_I_TANCAT
\^{\i}	"\\^\{\i}" \hat{\i}"	A_I_CIRCUM
\hat{\i}	"\\^\{\i}" \hat{\i}"	A_I_CIRCUM
\" {\i}	"\\\"{\i}" \ddot{\i}"	A_I_DIERESIS
\ddot{\i}	"\\\"{\i}" \ddot{\i}"	A_I_DIERESIS
\~{\i}	"\\~\{\i}" \tilde{\i}"	A_I_TILDE
\tilde{\i}	"\\~\{\i}" \tilde{\i}"	A_I_TILDE
\={\i}	"\\=\{\i}" \bar{\i}"	A_I_MACRON
\bar{\i}	"\\=\{\i}" \bar{\i}"	A_I_MACRON
\.\{\i}	"\\.\{\i}" \dot{\i}"	A_I_DOT
\dot{\i}	"\\.\{\i}" \dot{\i}"	A_I_DOT
\u{\i}	"\\u\{\i}" \breve{\i}"	A_I_BREVE
\breve{\i}	"\\u\{\i}" \breve{\i}"	A_I_BREVE
\' {\I}	"\\'\{\I}" \grave{\I}"	A_I_OBERT_M
\grave{\I}	"\\'\{\I}" \grave{\I}"	A_I_OBERT_M
\' {\I}	"\\'\{\I}" \acute{\I}"	A_I_TANCAT_M
\acute{\I}	"\\'\{\I}" \acute{\I}"	A_I_TANCAT_M
\^{\I}	"\\^\{\I}" \hat{\I}"	A_I_CIRCUM_M
\hat{\I}	"\\^\{\I}" \hat{\I}"	A_I_CIRCUM_M
\" {\I}	"\\\"{\I}" \ddot{\I}"	A_I_DIERESIS_M
\ddot{\I}	"\\\"{\I}" \ddot{\I}"	A_I_DIERESIS_M
\~{\I}	"\\~\{\I}" \tilde{\I}"	A_I_TILDE_M
\tilde{\I}	"\\~\{\I}" \tilde{\I}"	A_I_TILDE_M
\={\I}	"\\=\{\I}" \bar{\I}"	A_I_MACRON_M
\bar{\I}	"\\=\{\I}" \bar{\I}"	A_I_MACRON_M
\.\{\I}	"\\.\{\I}" \dot{\I}"	A_I_DOT_M
\dot{\I}	"\\.\{\I}" \dot{\I}"	A_I_DOT_M
\u{\I}	"\\u\{\I}" \breve{\I}"	A_I_BREVE_M
\breve{\I}	"\\u\{\I}" \breve{\I}"	A_I_BREVE_M

Taula 3.9: Macros d'accents i signes diacrítics sobre la lletra *i*

L ^A T _E X	Patró	Id. Token
\`{o}	"\\`{o}" "\grave{o}"	A_O_OBERT
\grave{o}	"\\`{o}" "\grave{o}"	A_O_OBERT
\`{o}	"\\' {o}" "\acute{o}"	A_O_TANCAT
\acute{o}	"\\' {o}" "\acute{o}"	A_O_TANCAT
\^{o}	"\\^{o}" "\hat{o}"	A_O_CIRCUM
\hat{o}	"\\^{o}" "\hat{o}"	A_O_CIRCUM
\" {o}	"\\\" {o}" "\ddot{o}"	A_O_DIERESIS
\ddot{o}	"\\\" {o}" "\ddot{o}"	A_O_DIERESIS
\~{o}	"\\~{o}" "\tilde{o}"	A_O_TILDE
\tilde{o}	"\\~{o}" "\tilde{o}"	A_O_TILDE
\={o}	"\\={o}" "\bar{o}"	A_O_MACRON
\bar{o}	"\\={o}" "\bar{o}"	A_O_MACRON
\u{o}	"\\u{o}" "\breve{o}"	A_O_BREVE
\breve{o}	"\\u{o}" "\breve{o}"	A_O_BREVE
\H{o}	"\\H{o}"	A_O_HUNGARIAN
\`{O}	"\\`{O}" "\grave{O}"	A_O_OBERT_M
\grave{O}	"\\`{O}" "\grave{O}"	A_O_OBERT_M
\`{O}	"\\' {O}" "\acute{O}"	A_O_TANCAT_M
\acute{O}	"\\' {O}" "\acute{O}"	A_O_TANCAT_M
\^{O}	"\\^{O}" "\hat{O}"	A_O_CIRCUM_M
\hat{O}	"\\^{O}" "\hat{O}"	A_O_CIRCUM_M
\" {O}	"\\\" {O}" "\ddot{O}"	A_O_DIERESIS_M
\ddot{O}	"\\\" {O}" "\ddot{O}"	A_O_DIERESIS_M
\~{O}	"\\~{O}" "\tilde{O}"	A_O_TILDE_M
\tilde{O}	"\\~{O}" "\tilde{O}"	A_O_TILDE_M
\={O}	"\\={O}" "\bar{O}"	A_O_MACRON_M
\bar{O}	"\\={O}" "\bar{O}"	A_O_MACRON_M
\u{O}	"\\u{O}" "\breve{O}"	A_O_BREVE_M
\breve{O}	"\\u{O}" "\breve{O}"	A_O_BREVE_M
\H{O}	"\\H{O}"	A_O_HUNGARIAN_M

Taula 3.10: Macros d'accents i signes diacrítics sobre la lletra o

L^AT_EX	Patró	Id. Token
<code>\`{u}</code>	<code>"\\`{u}"</code> <code>"\\grave{u}"</code>	A_U_OBERT
<code>\grave{u}</code>	<code>"\\`{u}"</code> <code>"\\grave{u}"</code>	A_U_OBERT
<code>\`{u}</code>	<code>"\\`{u}"</code> <code>"\\acute{u}"</code>	A_U_TANCAT
<code>\acute{u}</code>	<code>"\\`{u}"</code> <code>"\\acute{u}"</code>	A_U_TANCAT
<code>\^{u}</code>	<code>"\\^{u}"</code> <code>"\\hat{u}"</code>	A_U_CIRCUM
<code>\hat{u}</code>	<code>"\\^{u}"</code> <code>"\\hat{u}"</code>	A_U_CIRCUM
<code>\"u</code>	<code>"\\\"{u}"</code> <code>"\\ddot{u}"</code>	A_U_DIERESIS
<code>\ddot{u}</code>	<code>"\\\"{u}"</code> <code>"\\ddot{u}"</code>	A_U_DIERESIS
<code>\~{u}</code>	<code>"\\~{u}"</code> <code>"\\tilde{u}"</code>	A_U_TILDE
<code>\tilde{u}</code>	<code>"\\~{u}"</code> <code>"\\tilde{u}"</code>	A_U_TILDE
<code>\={u}</code>	<code>"\\={u}"</code> <code>"\\bar{u}"</code>	A_U_MACRON
<code>\bar{u}</code>	<code>"\\={u}"</code> <code>"\\bar{u}"</code>	A_U_MACRON
<code>\u{u}</code>	<code>"\\u{u}"</code> <code>"\\breve{u}"</code>	A_U_BREVE
<code>\breve{u}</code>	<code>"\\u{u}"</code> <code>"\\breve{u}"</code>	A_U_BREVE
<code>\H{u}</code>	<code>"\\H{u}"</code>	A_U_HUNGARIAN
<code>\r{u}</code>	<code>"\\r{u}"</code>	A_U_RING
<code>\`{U}</code>	<code>"\\`{U}"</code> <code>"\\grave{U}"</code>	A_U_OBERT_M
<code>\grave{U}</code>	<code>"\\`{U}"</code> <code>"\\grave{U}"</code>	A_U_OBERT_M
<code>\`{U}</code>	<code>"\\`{U}"</code> <code>"\\acute{U}"</code>	A_U_TANCAT_M
<code>\acute{U}</code>	<code>"\\`{U}"</code> <code>"\\acute{U}"</code>	A_U_TANCAT_M
<code>\^{U}</code>	<code>"\\^{U}"</code> <code>"\\hat{U}"</code>	A_U_CIRCUM_M
<code>\hat{U}</code>	<code>"\\^{U}"</code> <code>"\\hat{U}"</code>	A_U_CIRCUM_M
<code>\"U</code>	<code>"\\\"{U}"</code> <code>"\\ddot{U}"</code>	A_U_DIERESIS_M
<code>\ddot{U}</code>	<code>"\\\"{U}"</code> <code>"\\ddot{U}"</code>	A_U_DIERESIS_M
<code>\~{U}</code>	<code>"\\~{U}"</code> <code>"\\tilde{U}"</code>	A_U_TILDE_M
<code>\tilde{U}</code>	<code>"\\~{U}"</code> <code>"\\tilde{U}"</code>	A_U_TILDE_M
<code>\={U}</code>	<code>"\\={U}"</code> <code>"\\bar{U}"</code>	A_U_MACRON_M
<code>\bar{U}</code>	<code>"\\={U}"</code> <code>"\\bar{U}"</code>	A_U_MACRON_M
<code>\u{U}</code>	<code>"\\u{U}"</code> <code>"\\breve{U}"</code>	A_U_BREVE_M
<code>\breve{U}</code>	<code>"\\u{U}"</code> <code>"\\breve{U}"</code>	A_U_BREVE_M
<code>\H{U}</code>	<code>"\\H{U}"</code>	A_U_HUNGARIAN_M
<code>\r{U}</code>	<code>"\\r{U}"</code>	A_U_RING_M

Taula 3.11: Macros d'accents i signes diacrítics sobre la lletra *u*

L^AT_EX	Patró	Id. Token
<code>\c{c}</code>	<code>"\\c{c}"</code>	A_C_TRENCADA
<code>\c{C}</code>	<code>"\\c{C}"</code>	A_C_TRENCADA_M
<code>\~{n}</code>	<code>"\\~{n}"</code>	A_N_CEDILLA
<code>\~{N}</code>	<code>"\\~{N}"</code>	A_N_CEDILLA_M

Taula 3.12: Macros de lletres específiques

Capítol 4

Criteris, procediments i mètodes de traducció

A continuació s'expliquen els criteris, procediments i mètodes de traducció que utilitza el traductor alhora de traduir documents \LaTeX a HTML .

4.1 Conjunt de macros reconegudes pel traductor

Les macros permeten especificar una gran varietat d'aspectes que defineixen la forma en què es vol presentar el document, des de consideracions ortogràfiques i tipogràfiques, fins a l'estructura de tot el document. Hi ha prop d'un miler de macros estàndars, de les quals, se n'ha escollit un subconjunt bàsic. A continuació es mostren el conjunt de macros reconegudes pel traductor. No s'hi inclouen els accents ni els símbols, i tampoc s'hi onclouen les marques de composició de fragments en mode matemàtic.

Comandes `par`, `linebreak`, `newline`, `rule`, `includegraphics`, `epsfig`, `hline`, `documentclass`, `section`, `subsection`, `subsubsection`, `chapter`, `part`, `item`, `subitem`, `footnote`, `tableofcontents`, `listoffigures`, `listoftables`, `label`, `ref`, `hypertarget`, `hyperlink`, `href`, `paragraph`, `subparagraph`, `caption`, `title`, `author`, `and`, `date`, `thanks`, `maketitle`, `cite`, `bibitem`, `emph`, `textnormal`, `textit`, `textbf`, `underline`, `textsuperscript`, `texttt`

Declaracions `it`, `bf`, `em`, `tt`, `tiny`, `scriptsize`, `footnotesize`, `small`, `normalsize`, `large`, `Large`, `LARGE`, `huge`, `Huge`

Entorns `document`, `abstract`, `document`, `titlepage`, `enumerate`, `figure`, `table`, `tabular`, `equation`, `itemize`, `center`, `flushleft`, `flushright`, `tabbing`, `rawhtml`, `thebibliography`

4.2 Estructura d'un document L^AT_EX i HTML

Tot document L^AT_EX consisteix en dos parts ben diferenciades: el *preamble* i el *cos*.

El preàmbul és una col·lecció de macros que encapçalan el document, determinen el seu aspecte general i influeixen en el seu estil. En particular, una d'aquestes macros determina la classe de document que es tracta, com per exemple un article, un llibre, un reportatge d'investigació, etc...

El cos és el contingut del document. És a dir, el text del document conjuntament amb totes aquelles macros addicionals que resulten necessàries per a la seva composició. El preàmbul d'un document comença amb la macro

```
\documentclass[opcions]{classe}
```

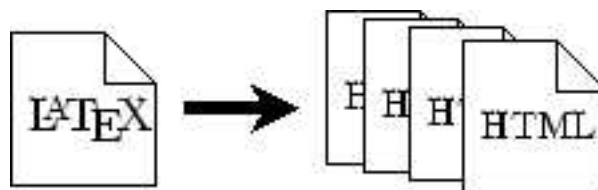
mentre que el cos comença amb la macro `\begin{document}` i acaba amb la macro `\end{document}`.

Tot allò comprès entre la macro `\documentclass` i la macro `\begin{document}` constitueix el preàmbul, és a dir, l'estructura d'un document L^AT_EX és la que s'indica a continuació:

```
\documentclass[opcions]{classe}
...
Preambul del document
...
\begin{document}
...
Cos del document
...
\end{document}
```

De forma similar, els documents HTML segueixen la següent estructura:

```
<HTML>
<HEAD>
...
Capçalera del document
...
</HEAD>
<BODY>
...
Cos del document
...
</BODY>
</HTML>
```



Per facilitar la lectura dels documents L^AT_EX resulta convenient dividir el cos del document en capítols, apartats i subapartats, d'acord amb l'estructura lògica del seu contingut. La divisió del contingut del document en unitats estructurals es fa mitjançant macros estructurals del L^AT_EX, les quals varien en funció de la classe de document de que es tracti. Algunes classes estàndars de documents L^AT_EX són: **article**, **report**, **book**, **letter** i **slides**.

Cada classe de document inclou macros propies. Degut a això, s'ha limitat la traducció únicament a documents amb macros de les classes **article**, **book** i **report**. El traductor ignora el tipus de classe del document que ha de traduir ja que per exemple, un document de tipus **letter** pot estar compostat per macros suportades pel traductor, i per tant, la traducció es pot realitzar sense cap mena de problema. D'aquesta manera, el traductor es limita a traduir tot allò que es capaç de traduir independentment de la classe de que es tracti. Les macros no suportades són ignorades pel traductor, per a més informació relacionada amb el comportament del traductor davant de macros no reconegudes es recomana fer una lectura a l'apartat A.5.

El traductor divideix el document en funció dels següents tipus d'apartats:

- Parts
- Capítols
- Seccions
- Subseccions
- Subsubseccions

Cada apartat es tradueix i es guarda en el seu corresponent fitxer HTML, així doncs, la traducció total d'un document consisteix a generar un conjunt de fitxers HTML, un per cada apartat.

4.3 Comptadors

El traductor contempla 4 comptadors dedicats a diferents elements comptables d'un document L^AT_EX: un comptador d'apartats, un comptador de figures, un comptador de taules i un comptador d'equacions. El comptador més destacat

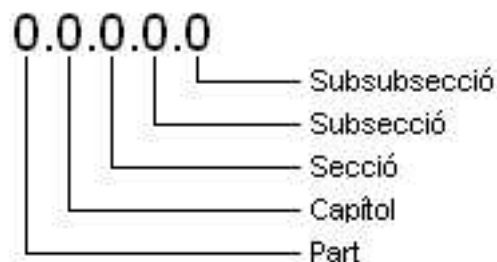


Figura 4.1: Comptador d'apartats

és el comptador d'apartats, que contempla fins a 5 nivells de compteig.

El comptador d'apartats s'ha implementat mitjançant la classe `comptador`, ja que el \LaTeX enumera automàticament els apartats en què es divideix el cos d'un document. El format del comptador és el que es mostra a la figura 4.1 i encapçula les següents principals operacions:

`void comptador::novaPart()` Incrementa el primer nivell del comptador i deixa la resta a zero.

`void comptador::nouCapitol()` Incrementa el segon nivell del comptador i deixa els tres últims a zero.

`void comptador::novaSeccio()` Incrementa el tercer nivell del comptador i deixa els dos últims a zero.

`void comptador::novaSubseccio()` Incrementa el quart nivell del comptador i deixa l'últim a zero.

`void comptador::novaSubsubseccio()` Incrementa el cinquè nivell del comptador.

`friend int comptador::CompCmp(comptador* c1, comptador* c2)` Compara *c1* i *c2*. Retorna 0 si *c1* o *c2* correspon a una subsecció de l'altre. Retorna 1 si *c1* correspon a una secció superior a *c2*. Retorna 2 si *c2* correspon a una secció superior a *c1*.

Quan finalitza un apartat i s'inicia la traducció del següent, s'actualitza el comptador d'apartats en funció de si el nou apartat és una part, un capítol, una secció, una subsecció o bé una subsubsecció.

La resta de comptadors destinats a altres propòsits són d'un únic nivell i per tant són definits com a *unsigned int*. En aquest sentit, els altres tres tipus de comptadors són iguals, excepte el comptador de figures i el comptador de taules que tenen una peculiaritat especial. El comptador d'equacions incrementa quan el traductor detecta la macro `\begin{equation}`, tanmateix els comptadors de figures i taules només incrementen quan el traductor detecta la macro *caption* dins de l'entorn *figure* o *table*. Els entorns *figure* i *table* que no contenen la macro *caption* no són comptabilitzats i per tant, tampoc apareixen en les llistes de figures i taules generades per les macros `\listoffigures` i `\listoftables`.

4.4 Pila d'entorns

La complexitat dels documents \LaTeX recau principalment en l'anidació d'entorns, és a dir, la composició d'entorns que es componen dins d'altres entorns. El problema de la traducció recau en aquelles situacions en les quals interessa conèixer determinades dades de l'entorn en curs i més tard, recuperar les dades de l'entorn del nivell o nivells inferiors.

Aquestes situacions es donen en els entorns comptables: **figure**, **table**, **equation** i el subconjunt de macros **part**, **chapter**, **section**, **subsection** i **subsubsection**. El traductor assigna a cada un d'aquests entorns, una etiqueta just a l'inici de cada un d'ells i evidentment, en tractar-se d'entorns comptables, és a dir, entorns amb una enumeració propia, també tenen assignat un valor de comptador. El valor d'aquestes etiquetes i comptadors és necessari alhora de crear referències creuades, ja que aquestes han de conèixer l'etiqueta i el valor del comptador dels entorns on s'han inserit les macros **label** (mirar l'apartat 4.10).

Per resoldre aquest problema, el traductor utilitza una pila d'entorns implementada mitjanant la classe **pilaEntorns** i la classe *entorn*.

Els atributs que encapsula la classe **entorn** són els següents:

`char* etiqueta;` Emmagatzema el valor d'una etiqueta.

`char* compt;` Emmagatzema el valor textual d'un comptador.

Els atributs i mètodes que encapsula la classe **pilaEntorns** són els següents:

`Llista<entorn*> l;` Llista que s'utilitza com una pila mitjanant les operacions estàndard `void Llista::inserirCap(T&), void Llista::eliminarCap();` i `void Llista::consultarCap(T&);`. Apila i desapila entorns.

`void pilaEntorns::apilarEntorn(char* etiqueta, char* comptador)` Apila l'entorn definit per l'etiqueta *etiqueta* i el comptador *comptador*.

```
void pilaEntorns::desapilarEntorn() Desapila l'últim entorn apilat.
```

```
void pilaEntorns::consultarEntornActual(char* etiqueta, char* comptador)
    etiqueta i comptador prenen per valor l'etiqueta i el comptador de l'ltim en-
    torn apilat(entorn en curs).
```

La figura 4.2 il·lustra amb claredat com funciona la pila d'entorns.

4.5 Accents i caràcters especials en mode text

La traducció de determinats accents i caràcters especials (símbols, signes de puntuació, lligadures, lletres gregues, etc...) és un dels principals problemes alhora de traduir un document. Els documents realitzats en \LaTeX han de respectar certes convencions pròpies del sistema \TeX , d'aquesta manera, els accents i altres caràcters especials s'inclouen seguint unes especificacions concretes. D'aquesta manera, en \LaTeX es poden obtenir nombrosos caràcters i símbols que HTML no és capaç de representar gràficament.

Per a la traducció s'ha escollit la codificació de caràcters *Unicode* (ISO-Latin-1 extés) suportada actualment per la versió HTML 4.0 , que entre d'altres coses, ha fet possible una extensió del nombre de caràcters de la codificació ISO-Latin-1, aquesta última usada fins a la versió d' HTML 3.2 i que actualment continua essent compatible amb navegadors que suporten HTML 4.0 o superior.

Per a la traducció es substitueix el caràcter especial en \LaTeX per la seva equivalent codificació en *Unicode*. Per exemple, el caràcter ü es representa en *Unicode* de la forma `ü`, d'aquesta manera, la traducció de la cadena *qüestió* a HTML és...

```
q&#252;esti&#243;
```

...on `ü` i `ó` són la codificació en *Unicode* de les lletres `ü` (`\{u}`) i `ó` (`\{o}`) respectivament.

Només un petit conjunt de totes les possibles lletres representables en \LaTeX tenen representació en HTML , això està en funció del traductor i de les codificacions de caràcters que suporti el navegador web. Els caràcters que no es poden representar en HTML no són suportats pel traductor, i aquest ho notifica amb un avís indicant la línia i la posició on s'ha trobat la macro desconeguda. Per a més informació relacionada amb el comportament del traductor davant de macros no reconegudes, es recomana fer una lectura a l'apartat A.5.

El procediment de traducció que aquí s'explica s'aplica exclusivament en els caràcters i símbols especials en mode text, és a dir, totes aquelles no incloses entre els símbols $\$$. El tractament de fragments matemàtics s'explica a l'apartat ??

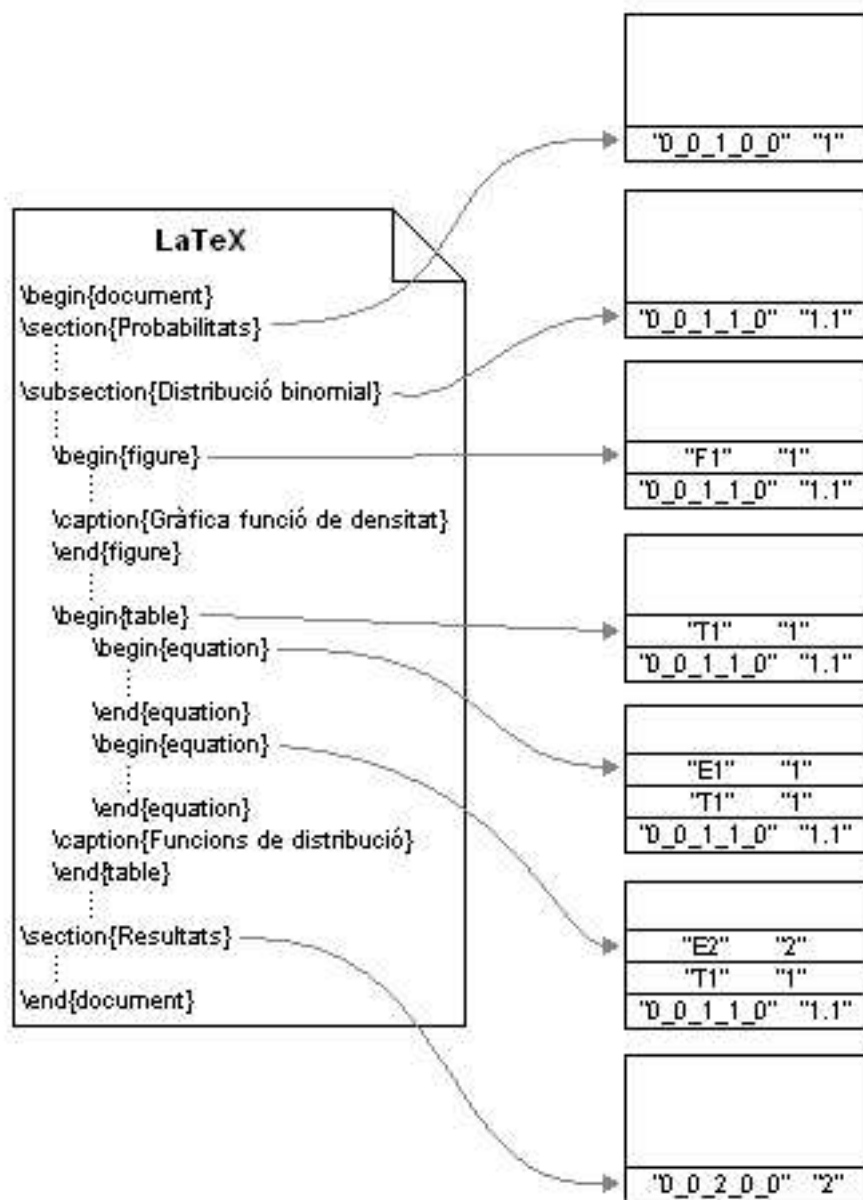


Figura 4.2: Funcionament de la pila d'entorns

Les següents taules mostren el conjunt de caràcters suportats pel traductor amb la seva corresponent macro en \LaTeX i codificació en *Unicode*.

- Taula 4.1: Símbols i caràcters especials
- Taula 4.2: Accents i signes diacrítics sobre la lletra *a*
- Taula 4.3: Accents i signes diacrítics sobre la lletra *e*
- Taula 4.4: Accents i signes diacrítics sobre la lletra *i*
- Taula 4.5: Accents i signes diacrítics sobre la lletra *o*
- Taula 4.6: Accents i signes diacrítics sobre la lletra *u*

4.6 Macros de funcionalitat específica

Aquí s’hi inclouen aquelles macros que tenen una funció concreta i que formen part de la pròpia construcció d’un text, formen part del sistema estructural de \LaTeX o simplement són un text ordinari.

4.6.1 Text ordinari

Es considera un text ordinari aquell fragment de text que la seva traducció a HTML és immediata, és a dir, textos sense macros, sense caràcters especials i sense marques estructurals del \LaTeX .

- una amplia gama de productos \leftarrow text de traducció immediata
- una amplia $\text{\emph{gama}}$ de productos \leftarrow text no ordinari
- la vida agr $\text{\'}{\text{\i}}$ cola \leftarrow text no ordinari

4.6.2 Final de paràgraf

Les macros \par i $\text{\}\}$ representen un final de paràgraf, i per tant, l’inici d’un nou paràgraf. La seva equivalència en HTML ve representada per la etiqueta \<P>

\LaTeX

...and 1988, were also imported from Asia. $\text{\}\}$ These guitars were generally considered to be good quality...

HTML

...and 1988, were also imported from Asia. \<P> These guitars were generally considered to be good quality...

Símbol	L ^A T _E X	Unicode
\$	\\$ o \textdollar	$
&	\&	&
%	\%	%
-	_ o \textunderscore	_
{	\{ o \textbraceleft	{
}	\} o \textbraceright	}
<	\textless	<
>	\textgreater	>
≠	\neq	≠
≤	\leq	≤
≥	\geq	≥
≈	\approx	≈
≡	\equiv	≡
∞	\infty	∞
⊃	\cap	∩
∫	\int	∫
∑	\sum	∑
∏	\prod	∏
△	\triangle	∆
∂	\partial	∂
\	\textbackslash	\
	\textbar	|
•	\textbullet	•
‡	\textdaggerdbl	‡
†	\textdagger	†
¶	\textparagraph	¶
§	\textsection	§
©	\textcopyright	©
ˆ	\textasciicircum	^
˜	\textasciitilde	˜
®	\textregistered	®
™	\texttrademark	™
ª	\textordfeminine	ª
º	\textordmasculine	º
ß	\ss	ß

Taula 4.1: Símbols i caràcters especials

Lletra	L ^A T _E X	Unicode
à	\`{a} o \grave{a}	à
á	\'{a} o \acute{a}	á
â	\^{a} o \hat{a}	â
ä	\">{a} o \ddot{a}	ä
ã	\~{a} o \tilde{a}	ã
ā	\={a} o \bar{a}	ā
ă	\u{a} o \breve{a}	ă
å	\r{a}	å
æ	\ae	æ
À	\`{A} o \grave{A}	À
Á	\'{A} o \acute{A}	Á
Â	\^{A} o \hat{A}	Â
Ä	\">{A} o \ddot{A}	Ä
Ã	\~{A} o \tilde{A}	Ã
Ā	\={A} o \bar{A}	Ā
Ă	\u{A} o \breve{A}	Ă
Å	\r{A}	Å
Æ	\AE	Æ

Taula 4.2: Macros d'accents i signes diacrítics sobre la lletra *a*

Lletra	L ^A T _E X	Unicode
è	\`{e} o \grave{e}	è
é	\'{e} o \acute{e}	é
ê	\^{e} o \hat{e}	ê
ë	\">{e} o \ddot{e}	ë
ē	\={e} o \bar{e}	ē
è	\.{e} o \dot{e}	ė
ě	\u{e} o \breve{e}	ĕ
ě	\v{e} o \check{e}	ě
È	\`{E} o \grave{E}	È
É	\'{E} o \acute{E}	É
Ê	\^{E} o \hat{E}	Ê
Ë	\">{E} o \ddot{E}	Ë
Ē	\={E} o \bar{E}	Ē
Ė	\.{E} o \dot{E}	Ė
Ě	\u{E} o \breve{E}	Ĕ
Ě	\v{E} o \check{E}	Ě

Taula 4.3: Macros d'accents i signes diacrítics sobre la lletra *e*

Lletra	L ^A T _E X	Unicode
ì	\`{\i} o \grave{\i}	ì
í	\'{\i} o \acute{\i}	í
î	\^{\i} o \hat{\i}	î
ï	\" {\i} o \ddot{\i}	ï
ĩ	\~{\i} o \tilde{\i}	ĩ
ī	\={\i} o \bar{\i}	ī
ï	\. {\i} o \dot{\i}	i
ï	\u{\i} o \breve{\i}	ĭ
Ì	\`{I} o \grave{I}	Ì
Í	\'{I} o \acute{I}	Í
Î	\^{I} o \hat{I}	Î
Ï	\" {I} o \ddot{I}	Ï
Ĩ	\~{I} o \tilde{I}	Ĩ
ī	\={I} o \bar{I}	Ī
İ	\. {I} o \dot{I}	İ
Ĭ	\u{I} o \breve{I}	Ĭ

Taula 4.4: Macros d'accents i signes diacrítics sobre la lletra *i*

Lletra	L ^A T _E X	Unicode
ò	\`{\o} o \grave{o}	ò
ó	\'{\o} o \acute{o}	ó
ô	\^{\o} o \hat{o}	ô
ö	\" {\o} o \ddot{o}	ö
õ	\~{\o} o \tilde{o}	õ
ō	\={\o} o \bar{o}	ō
ö	\u{o} o \breve{o}	ŏ
ö	\H{o}	ő
Ò	\`{O} o \grave{O}	Ò
Ó	\'{O} o \acute{O}	Ó
Ô	\^{O} o \hat{O}	Ô
Ö	\" {O} o \ddot{O}	Ö
Õ	\~{O} o \tilde{O}	Õ
Ō	\={O} o \bar{O}	Ō
Ö	\u{O} o \breve{O}	Ŏ
Ö	\H{O}	Ő

Taula 4.5: Macros d'accents i signes diacrítics sobre la lletra *o*

Lletra	L ^A T _E X	Unicode
ù	\‘{u} o \grave{u}	ù
ú	\’{u} o \acute{u}	ú
û	\^{}{u} o \hat{u}	û
ü	\”{u} o \ddot{u}	ü
ũ	\~{}{u} o \tilde{u}	ũ
ū	\={u} o \bar{u}	ū
ŭ	\u{u} o \breve{u}	ŭ
Ũ	\H{u}	ű
Ů	\r{u}	ů
Ù	\‘{U} o \grave{U}	Ù
Ú	\’{U} o \acute{U}	Ú
Û	\^{}{U} o \hat{U}	Û
Ü	\”{U} o \ddot{U}	Ü
Ũ	\~{}{U} o \tilde{U}	Ũ
Ū	\={U} o \bar{U}	Ū
Ŭ	\u{U} o \breve{U}	Ŭ
Ũ	\H{U}	Ű
Ů	\r{U}	Ů

Taula 4.6: Macros d’accents i signes diacrítics sobre la lletra *u*

Representació

...and 1988, were also imported from Asia.

These guitars were generally considered to be good quality...

4.6.3 Trencament de línia

Les macros `\linebreak` i `\newline` representen un trencament de línia. La seva equivalència en `HTML` ve representada per la etiqueta `
`

L^AT_EX

...These `\newline` guitars were generally considered to be good quality...

HTML

...These `
` guitars were generally considered to be good quality...

Representació

...These
guitars were generally considered to be good quality...

4.6.4 Línia horitzontal

La macro `\rule` representa gràficament una línia horitzontal en funció de la seva llargada i gruix. La seva equivalència a HTML ve representada per la etiqueta `<HR>` seguint el següent esquema d'ús:

- **L^AT_EX**: `\rule{longitud}{gruix}`
- **HTML**: `<hr width="longitud"size="gruix">`

L'única unitat de mesura que utilitza el llenguatge HTML és el píxel, d'aquesta manera el traductor ignora el tipus d'unitats de mesura utilitzat en el document, de tal manera que el valor numèric dels camps *longitud* i *gruix* serà tractat com a píxels sigui quina sigui la unitat de mesura que hi hagi indicada.

- **L^AT_EX**: `\rule{100mm}{1cm} \rule{10mm}{2mm}`
- **HTML**: `<hr width="100"size="1"> <hr width="10"size="2">`

4.6.5 Comentaris

Els documents L^AT_EX permeten la inclusió de comentaris mitjançant l'ús del caràcter `%` seguit del comentari. El comentari finalitza quan s'acaba la línia, és a dir, quan es troba el caràcter `\n`

```
\emph{minority artist creation of
% Això és un comentari i no afecta al document
which students need to be proud}
```

Els comentaris són ignorats pel traductor, donat que no tenen efecte en el document.

4.7 Marques de format de text

Les marques de format de text permeten definir el format del text del document. El traductor suporta els següents formats:

- Emfatitzat
- Itàlica
- Mecnigràfic
- Negreta
- Subratllat

- Superscript
- Text normal
- Anidació de formats de text iguals

4.7.1 Text emfatitzat

La marca `\emph` permet emfatitzar un text. La seva equivalència a HTML ve representada per la etiqueta `` seguint el següent esquema:

- **L^AT_EX**: `\emph{Die Another Day}`
- **HTML**: `Die Another Day`

4.7.2 Text itàlica

La marca `\textit` posa en itàlica un fragment de text. La seva equivalència a HTML ve representada per la etiqueta `<I>` seguint el següent esquema:

- **L^AT_EX**: `\textit{Die Another Day}`
- **HTML**: `<I>Die Another Day</I>`

4.7.3 Text mecanogràfic

La marca `\texttt` posa en format mecanogràfic un fragment de text. La seva equivalència a HTML ve representada per la etiqueta `<TT>` seguint el següent esquema:

- **L^AT_EX**: `\texttt{Die Another Day}`
- **HTML**: `<TT>Die Another Day</TT>`

4.7.4 Text negreta

La marca `\textbf` posa en negreta un fragment de text. La seva equivalència a HTML ve representada per la etiqueta `` seguint el següent esquema:

- **L^AT_EX**: `\textbf{Die Another Day}`
- **HTML**: `Die Another Day`

4.7.5 Text subratllat

La marca `\underline` subratlla un fragment de text. La seva equivalència a HTML ve representada per la etiqueta `<U>` seguint el següent esquema:

- **L^AT_EX**: `\underline{Die Another Day}`
- **HTML**: `<U>Die Another Day</U>`

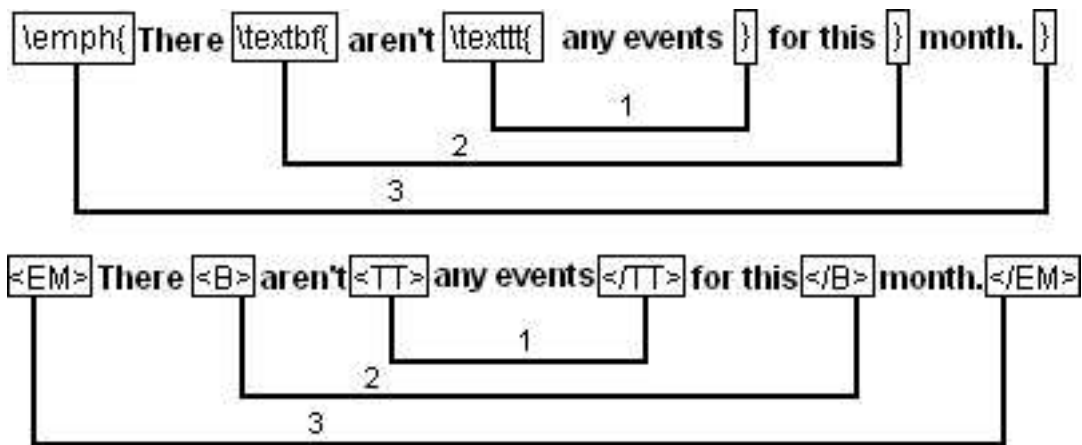


Figura 4.3: Formats de text sobreposats

4.7.6 Text superscript

La marca `\textsuperscript` posa un text com a superíndex. La seva equivalència a HTML ve representada per la etiqueta `<SUP>` seguint el següent esquema:

- **L^AT_EX**: `Die Another Day`
- **HTML**: `^{Die Another Day}`

4.7.7 Text normal

La marca `\textnormal` reestableix el format del text al format normal. Aquesta macro no té equivalent en HTML i per tant, s'ha d'optar per usar una alternativa.

El procés de traducció funciona recursivament en aquelles macros que actuen en forma de contenidor, és a dir, aquelles macros que tenen efecte dins d'un entorn tancat per claus el qual pot contenir altres entorns anidats. Les macros de format de text tenen aquest tipus de comportament, això vol dir que es poden sobreposar diferents formats de text en un mateix fragment de text. La figura 4.3 en mostra un exemple.

Reestablir un text al seu format normal implica conèixer els formats actius del text en curs. D'aquesta manera el problema es resol desactivant els formats a l'inici i tornant-los a activar just quan la macro `\textnormal` finalitza. Per a fer possible això s'utilitzen *flags*, un per cada un dels formats de text. La figura 4.4 en mostra un exemple.

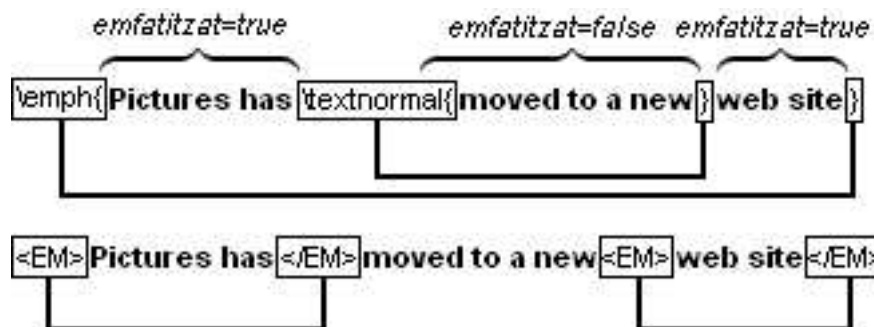


Figura 4.4: Desactivació de formats de text

4.7.8 Anidació de formats de text iguals

L'anidació de marques iguals de format de text segueix un comportament especial en \LaTeX . Per exemple, posar en format emfatitzat un text que ja està en format emfatitzat produeix el següent resultat:

```
\emph{To understand \emph{and \emph{protect} our} home planet}
      To understand and protect our home planet
```

El problema es resol invertint el valor del *flag* del format de text representat per la macro en qüestió. Així doncs, s'aconsegueix simular l'efecte produït per l'anidació de formats de text iguals. La figura 4.5 en mostra un exemple.

4.8 Subconjunt `maketitle`, `title`, `author`, `date`, `thanks` i `and`

Aquest subconjunt de macros permeten definir un títol complet per al document a la zona del preàmbul. Aquest títol inclou el propi títol, l'autor o autors i la data del propi document.

Es declara el títol i la informació de l'autor mitjançant la macro `\title` i `\author`, cada un accepta un simple argument. Múltiples autors poden ser llistats a l'argument de la macro `\author`, separats per la macro `\and`.

```
\title{The Language Evolution}
\author{Marc D. Hauser \and C.Chomsky}
```

La macro `\date` permet incloure la data del document. Si aquesta macro no s'inclou es posarà la data actual en el moment que es tradueix.

La macro `\thanks` permet introduir agraïments, direccions dels autors i altres textos que es componen com a notes a peu de pàgina.

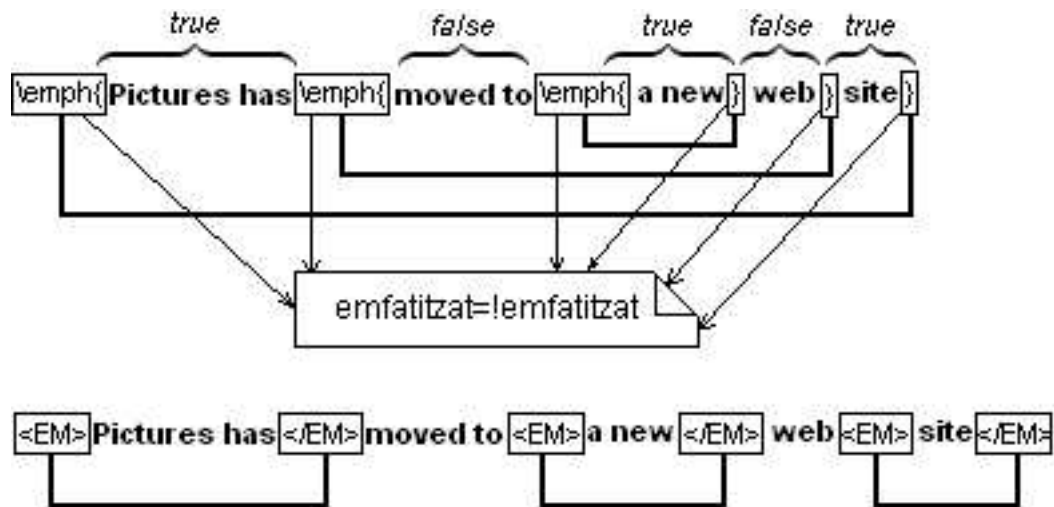


Figura 4.5: Formats de text iguals anidats

```
\title{The Language Evolution \thanks{It is the original title}}
\author{Marc D. Hauser\thanks{From Germany} \and C.Chomsky}
```

Un cop declarades totes aquestes dades al preàmbul, s'ha d'incloure la macro `\maketitle` just després de la macro `\begin{document}`, és a dir, ha d'estar afegit just al principi del cos del document.

La traducció del títol a HTML es realitza mitjançant l'etiqueta `<H1>` juntament amb alineació de text centrat de la següent manera:

```
\title{The Language Evolution}
<h1 align=center>The Language Evolution</h1>
```

La traducció de la secció de l'autor i la data a HTML es realitza mitjançant l'etiqueta `<H3>` juntament amb alineació de text centrat. La macro `\and` es tradueix per dos símbols seguits ` `; de la següent manera:

L^AT_EX

```
\author{Marc D. Hauser \and C.Chomsky}
\date{November 2002}
```

HTML

```
<h3 align=center>Marc D. Hauser &nbsp;&nbsp;  C.Chomsky</h3>
<h3 align=center>November 2002</h3>
```


The Language Evolution

Marc D. Hauser C. Chomsky

November 2002

Figura 4.6: Representació gràfica d'un títol en un navegador web

La figura 4.6 mostra la representació gràfica dels exemples anteriors un cop carregat el codi HTML en un navegador web corrent.

La macro `\thanks` es tradueix exactament igual que si es tractés d'una notació a peu de pàgina, igual que en el cas de la macro `\footnote`, però les numeracions no es realitzen amb números, sinó mitjançant els següents símbols: *, †, ‡, §, ¶, ||, **, †† i †††

4.9 Notes a peu de pàgina. Macro footnote

La macro `\footnote` permet afegir notes numerades a peu de pàgina. HTML és un llenguatge dinàmic que permet accedir a determinades zones d'un document web mitjançant enllaços a referències, d'aquesta manera, dotar amb links els peus de pàgina i les citacions a peu de pàgina millora notablement les possibilitats dels documents finals generats.

En el transcurs de la traducció d'un document, quan es detecta una citació a una nota a peu de pàgina, és a dir, quan es localitza una marca `\footnote`, s'incrementa el comptador de notes a peu de pàgina, s'emmagatzema la nota a memòria juntament amb el valor del comptador i s'assigna un nom a la nota i un nom al punt on s'ha localitzat la citació. Un cop fet tot això, la traducció consisteix en posar el valor del comptador en aquella part del document HTML juntament amb el *link* corresponent a la seva nota de peu associada.

Quan s'arriba al final de la traducció de l'apartat en curs, s'escriuen totes les notes a peu de pàgina memoritzades juntament amb els seus enllaços associats i noms. Amb això s'aconsegueix que fent *click* sobre una citació a una nota a peu de pàgina s'accedeix directament a la nota i viceversa. L'exemple de la figura 4.7 mostra el resultat de les notes al peu de pàgina en un navegador web.

El fet de decidir posar les notes de peu de pàgina al final de cada apartat es degut a que en HTML no té sentit parlar de pàgines ja que el document no

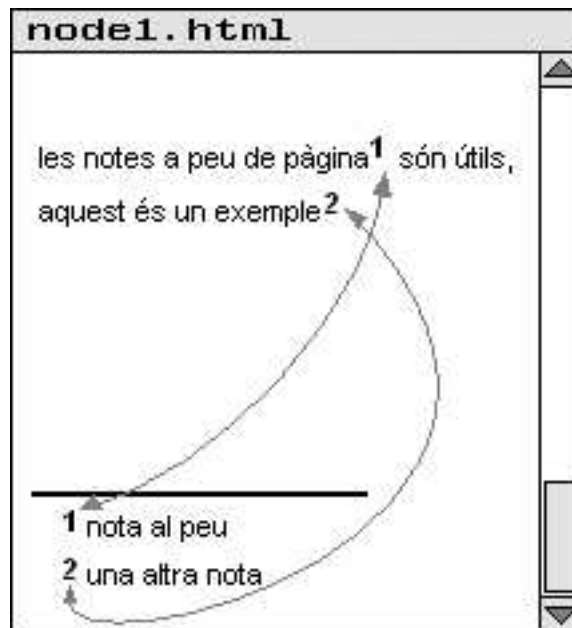


Figura 4.7: Notes al peu de pàgina en un navegador web

s'estructura en pàgines sinó en apartats estructurals. Per a més informació es recomana fer una lectura a l'apartat 4.2.

L^AT_EX

les notes a peu de pàgina\footnote{nota al peu} són útils, aquest és un exemple\footnote{una altra nota}

HTML

```
<HTML>
:
:
les notes a peu de pàgina<a href=node1.html#NPPN1 name=NPPF1><sup>1</sup></a>són
útils, aquest és un exemple<a href=node1.html#NPPN2 name=NPPF2><sup>2</sup></a>
:
<p><hr align=left width=175 size=1><small>
&nbsp;&nbsp;&nbsp;<a name=NPPN1 href=node1.html#NPPF1><sup>1</sup></a>nota
al peu<BR>
&nbsp;&nbsp;&nbsp;<a name=NPPN2 href=node1.html#NPPF2><sup>2</sup></a>una
altra nota<BR>
</small></p>
</HTML>
```

Per a la traducció de les notes a peu de pàgina s'ha implementat la classe `footnotes` de la que cal destacar els següents atributs i operacions encapsulades:

`Llista<nota*>lnotes;` És la llista de notes que emmagatzema totes les notes de l'apartat en curs.

`Llista<nota*>lgracies;` És la llista d'agraïments que emmagatzema els agraïments definits amb la macro `\thanks`.

`int compt;` És el comptador de notes.

`int compt2;` És el comptador d'agraïments.

`char* footnotes::obtenirEtiquetaFont(bool tipus)` Genera i retorna una etiqueta per a una citació en funció de si es demana una etiqueta per a una nota o una etiqueta per a un agraïment. Les etiquetes per a les notes es generen seguint el patró `NPPF*` on `*` és el valor del comptador de notes. Les etiquetes per als agraïments es generen seguint el patró `GPPF*` on `*` és el valor del comptador d'agraïments.

`char* footnotes::obtenirEtiquetaNota(bool tipus)` Genera i retorna una etiqueta per a una nota al peu de pàgina en funció de si es demana una etiqueta per a una nota o una etiqueta per a un agraïment. Les etiquetes per a les notes es generen seguint el patró `NPPN*` on `*` és el valor del comptador de notes. Les etiquetes per als agraïments es generen seguint el patró `GPPN*` on `*` és el valor del comptador d'agraïments.

`void footnotes::guardarNota(char* nota, char*& num, char*& etiqueta_font, char*& etiqueta_nota)` Afegeix una nota a la llista de notes, genera i retorna per paràmetre les etiquetes que s'han d'incloure en el punt de citació i el valor del comptador. Finalment incrementa el comptador de notes.

`void footnotes::guardarGracies(char* agraïment, char*& num, char*& etiqueta_font, char*& etiqueta_nota)` Afegeix un agraïment a la llista d'agraïments, genera i retorna per paràmetre les etiquetes que s'han d'incloure en el punt de citació i el valor del comptador. Finalment incrementa el comptador d'agraïments.

`void footnotes::buidar()` Buida la llista de notes i la llista d'agraïments i reinicia novament els comptadors a zero.

`char* footnotes::intToAsc(int numero, bool tipus)` Aquesta funció transforma l'enter del primer paràmetre en una cadena de caràcters en funció del tipus de transformació que es vol realitzar. Si es tracta d'una nota a peu de pàgina la transformació s'efectua amb normalitat, si es tracta d'un agraïment es realitza la transformació seguint els següents criteris:

- Retorna `*` si `numero == 1`
- Retorna `†` si `numero == 2`
- Retorna `‡` si `numero == 3`
- Retorna `§` si `numero == 4`
- Retorna `¶` si `numero == 5`
- Retorna `||` si `numero == 6`
- Retorna `**` si `numero == 7`
- Retorna `††` si `numero == 8`
- Retorna `‡‡` si `numero == 9`
- Altrament retorna un espai en blanc

`char* footnotes::crearPeuPagina()` Genera i retorna el peu de pàgina en HTML corresponent a l'apartat en curs. Finalment buida les llistes de notes i d'agraïments per deixar-les preparades per al pròxim apartat a traduir.

4.10 Referències creuades. Subconjunt `label` i `ref`

Les macros `\label` i `\ref` serveixen per incloure referències a punts concrets del document \LaTeX . Una referència creuada inclou la identificació numèrica d'un apartat o entorn comptable del document, com pot ser una figura, una (sub)(sub)secció, una part, un capítol, una taula o una equació.

Quan es compila un document \LaTeX , el compilador reemplaça en el document tota macro inclosa en l'original electrònic de la forma `\ref{etiqueta}` per la referència a la part corresponent del document, com ara el número d'apartat, figura, taula o teorema on s'ha inclòs la macro `\label{etiqueta}`. El següent exemple detalla el funcionament de les referències creuades en \LaTeX i la figura 4.8 en mostra la seva representació gràfica en un entorn HTML.

```
 $\text{\LaTeX}$ 
\section{Resum de l'article}
\label{resum}
```

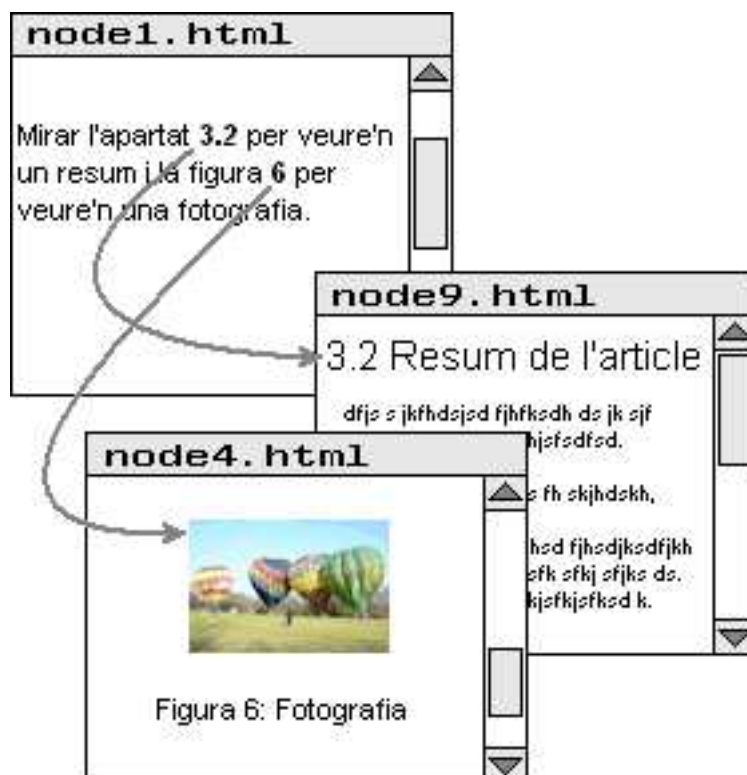


Figura 4.8: Referències creuades en un entorn HTML

```

:
\begin{figure}
\caption{Fotografia}
\label{foto}
\end{figure}
:
Mirar l'apartat \ref{resum} per veure'n un resum i la figura \ref{foto}
per veure'n una fotografia.

```

Representació

Mirar l'apartat 3.2 per veure'n un resum i la figura 6 per veure'n una fotografia.

La traducció de les referències creuades es porta a terme realitzant dos escombrades al document. La primera escombrada s'efectua durant la traducció normal del document i realitza dos tasques:

- Es guarden en una taula de hash totes les etiquetes definides per les macros `\label{etiqueta}` juntament amb l'etiqueta de l'entorn on es troben, el nom del fitxer HTML que s'està generant i el valor del comptador de l'entorn on es troben. A continuació es reemplaça cada una d'aquestes macros per `` on `*` és el valor del comptador d'etiquetes, inicialment amb valor 1. Finalment, s'incrementa el comptador.
- Es reemplacen les macros `\ref{etiqueta}` per la cadena `{Retiqueta}` on `etiqueta` és la mateixa etiqueta definida per l'usuari amb la macro `\ref`. Aquesta cadena, és d'ús reservat pel traductor.

La segona escombrada es realitza un cop finalitzada la traducció del document, aquest cop es fa sobre els documents HTML generats. Es cerquen les cadenes que compleixen el patró `{Retiqueta}` on `etiqueta` és una etiqueta definida per l'usuari. Aquestes són les cadenes generades durant la primera escombrada. Poden haver-hi altres cadenes que compleixen aquest patró i no han estat generades pel traductor?, la resposta és no, ja que els caràcters `{ i }` es tradueixen mitjançant la seva codificació en *Unicode*, a més a més, el llenguatge de marcatge HTML no utilitza aquests símbols en la seva sintaxi. En definitiva, els caràcters `{ i }` són caràcters reservats d'ús exclusiu del traductor durant el procés de traducció. Cada vegada que es localitza una d'aquestes cadenes es comprova si l'etiqueta existeix a la taula d'etiquetes.

Si hi existeix es tradueix per la següent estructura HTML:

```
<a href=fitxer.html#etiqueta_entorn>num_compt</a>
```

on `fitxer.html`, `etiqueta_entorn` i `num_compt` són camps del registre cercat a la taula d'etiquetes. Això permetrà que en el document HTML final en clicar damunt de la referència accedim directament a la part referenciada.

Si l'etiqueta no existeix a la taula, aleshores vol dir que s'està fent una referència a una etiqueta que no existeix al document i el traductor ho notifica a l'usuari amb un avís i continua el procés amb normalitat. La figura 4.9 mostra amb detall la traducció de referències creuades.

Els registres de la taula tenen el següent format:

char* etiqueta	char* etiqueta_entorn	char* num_compt	char* fitxer
----------------	-----------------------	-----------------	--------------

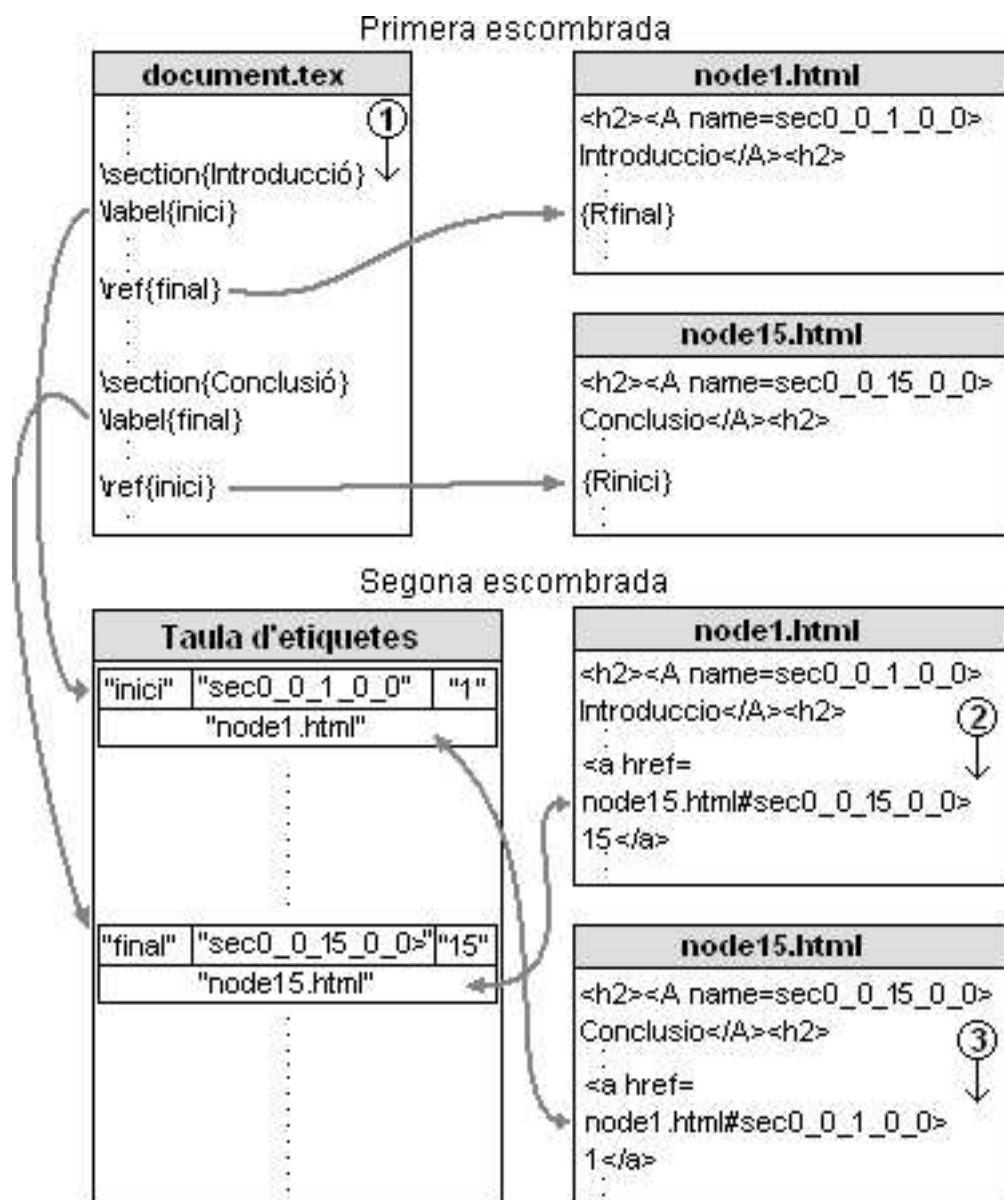


Figura 4.9: Traducció de referències creuades

4.11 Subconjunt part, chapter, section, subsection i subsubsection

En aquest apartat s'explica principalment com s'estructura el procés de traducció. Una de les tasques principals del traductor, a part de realitzar la traducció dels elements textuais d'un document, és la de fragmentar la informació en fitxers HTML, també anomenats *nodes*. La descomposició del document en varis fitxers HTML és necessària si volem aconseguir interactivitat entre els diferents apartats.

La traducció d'un document és seqüencial des de el punt de vista de que la informació es va traduint a mesura que es va llegint del fitxer d'origen, però no és seqüencial en el sentit de que la informació està classificada i enllaçada seguint uns mecanismes de referències creuades. És obvi que si volem obtenir un document interactiu(hipertextual) no és suficient una única traducció seqüencial del document.

El traductor manté en memòria, durant tot el procés, cadascun dels apartats traduïts en una estructura d'arbre jeràrquica. Cada node de l'arbre correspon a un apartat estructural del document i en aquests s'hi emmagatzema el codi HTML generat d'aquell apartat, el nom de fitxer que té assignat, el títol de l'apartat, un comptador d'apartats amb el valor de l'apartat en qüestió i tres punters a altres nodes. Aquests tres punters són: un punter a l'apartat anterior, un punter a l'apartat superior(pare) i un punter al següent apartat.

Per fer possible aquesta estructuració s'ha implementat la classe **aseccions** que s'encarrega, d'entre altres coses, de la gestió i manteniment de cadascun dels apartats que componen el document. Les tasques principals d'aquesta classe són:

1. Proporcionar la gestió i manteniment de cadascun dels apartats traduïts durant el transcurs de la traducció.
2. Materialitzar els nodes de l'arbre en els seus corresponents fitxers assignats un cop finalitzada la traducció.
3. Crear els panells de navegació de cada node de l'arbre.
4. Gestionar i mantenir els enllaços entre els diferents nodes.
5. Crear la taula de continguts.
6. Completar les operacions pendents
7. Crear el fitxer dels marcs i el fitxer del marc esquerra.
8. Inserir les capçaleres i els peus al codi HTML dels nodes.
9. Mantenir un comptador d'apartats.

A continuació es detallen les operacions principals que encapçala la classe:

`void aseccions::novaSeccio(int tipus, char* titol, char* cos)` Insereix un nou node a l'arbre. El nou node s'insereix ordenadament en funció de l'apartat que li correspon i se li assigna un nom de fitxer del tipus `node?.html`. S'enllaça o apunta el node anteriorment inserit amb l'actual i l'actual s'enllaça amb el node pare. La traducció seqüencial del document garanteix que els nodes sempre s'insereixen ordenadament, de tal manera que el següent node a inserir sempre és el següent apartat del document. Finalment actualitza el comptador d'apartats.

`void aseccions::CrearTaulaContinguts()` Crea el codi HTML corresponent a la representació de la taula de continguts i el guarda en un espai de memòria reservat. Aquest procés es porta a terme fent un recorregut ordenat a través de l'arbre.

`char* aseccions::crearCapcalera()` Retorna un espai de memòria que conté la capçalera que s'ha d'afegir a l'inici del codi HTML de cada node. Aquesta capçalera inclou les marques `HEAD` i `TITLE`.

`char* aseccions::crearIniciBody()` Retorna un espai de memòria que conté el format de les marques `BODY` que s'hauran d'afegir després de la capçalera del codi HTML de cada node. Aquest inclou les marques `BGCOLOR`, `LINK`, `VLINK` i `ALINK`.

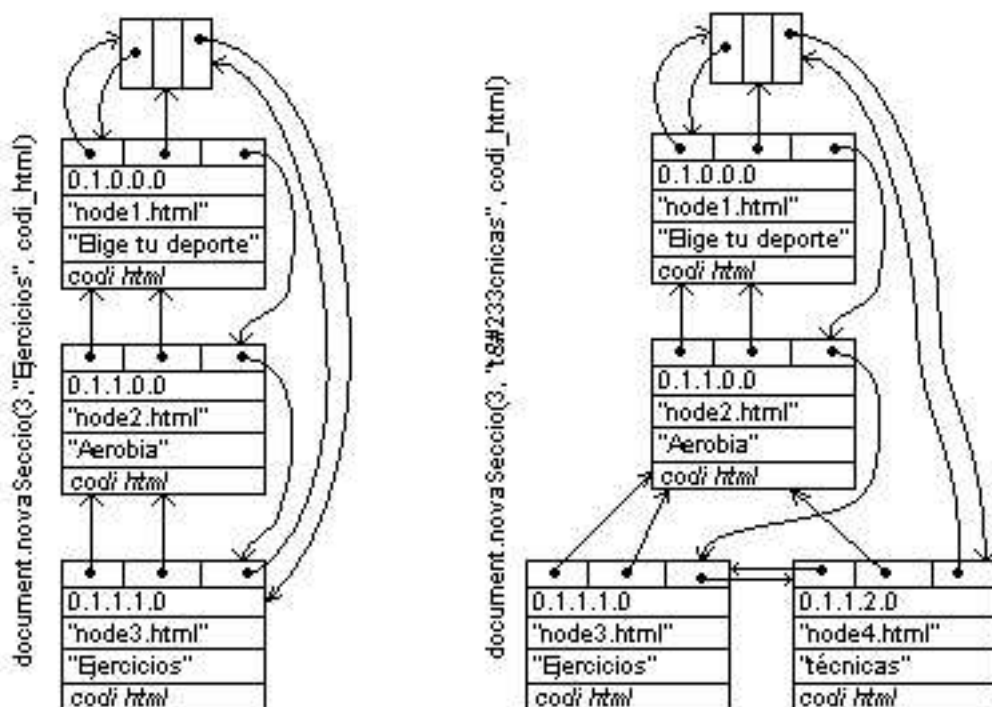
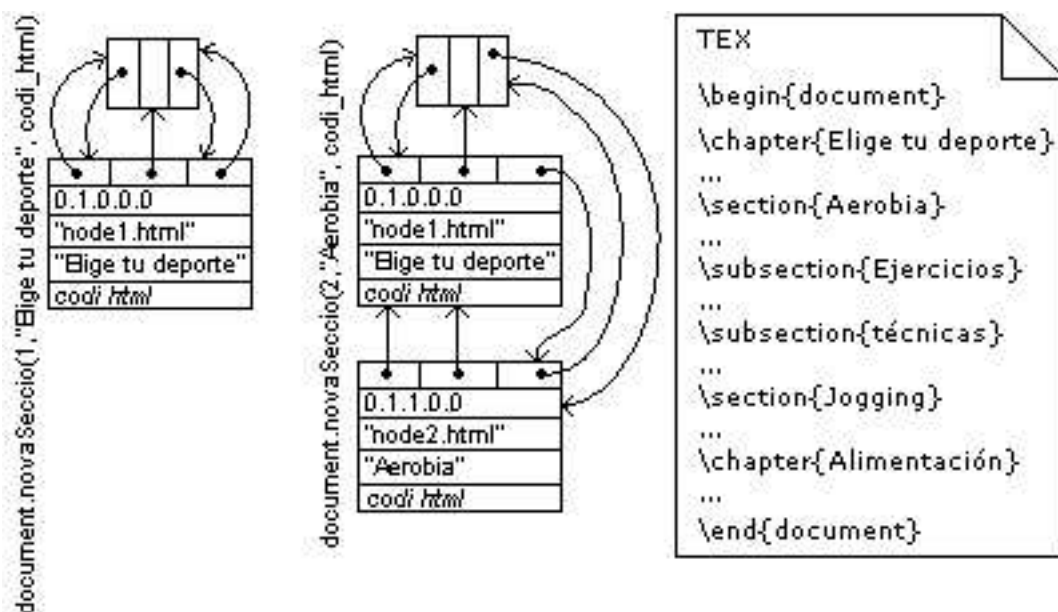
`void aseccions::crearFitxer(char* nom_fitxer, char* codi_html)` Crea un fitxer de text que té per nom `nom_fitxer` i el seu contingut és `codi_html`.

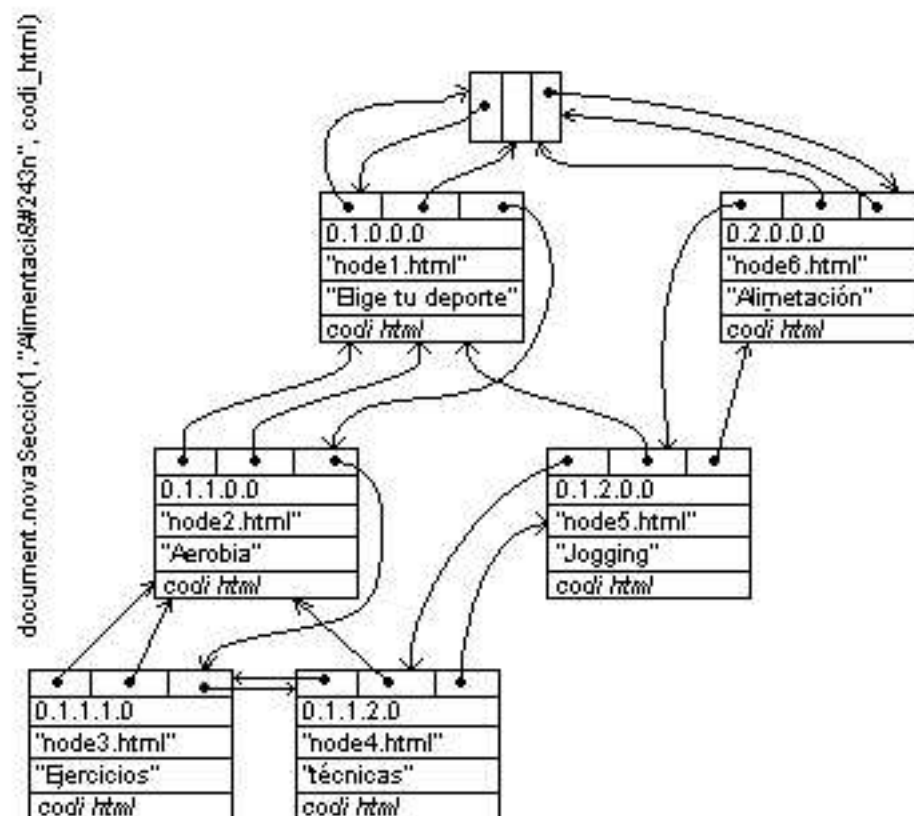
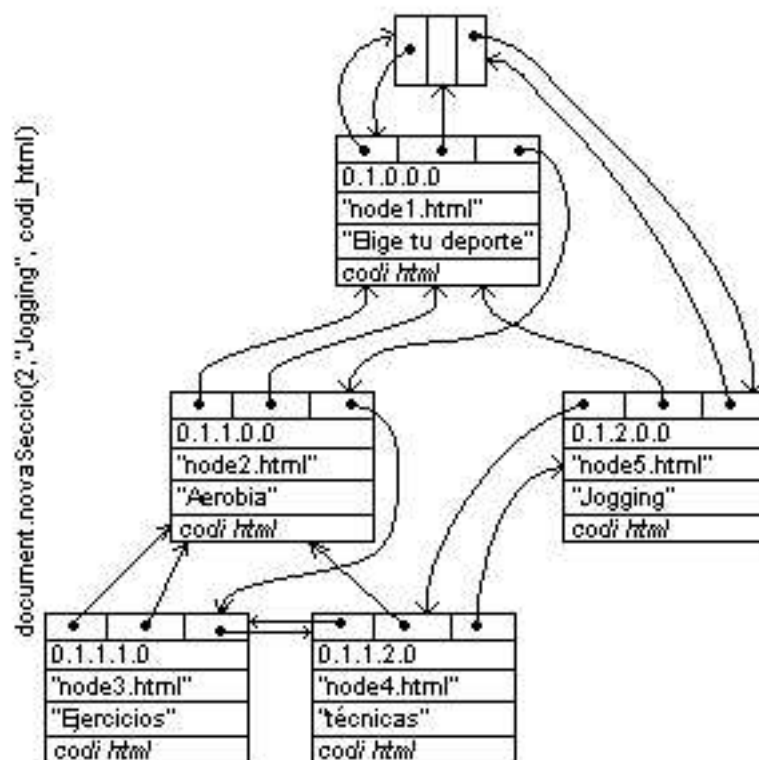
`void aseccions::crearFitxerTaulaContinguts()` Crea el fitxer de text `continguts.html` que conté la taula de continguts del document en HTML.

`void aseccions::crearFitxerFrames()` Crea el fitxer de text `index.html` que conté el codi en HTML per generar els marcs de la web. El marc esquerra s'associa amb el fitxer `continguts.html` i el marc dret s'associa amb el fitxer `node1.html`.

`char* aseccions::crearPanellNavegacio(seccio* s)` Retorna un punter que apunta a un espai de memòria que conté el panell de navegació associat a l'apartat o node `s`.

`void aseccions::crearWeb(crossRef* etiquetes)` S'encarrega de completar les operacions pendents i finalment materialitza els nodes de l'arbre en els seus corresponents fitxers. Per completar les operacions pendents, es realitza un recorregut ordenat a través de cada node de l'arbre i per a cada node es completen les referències creuades, les citacions bibliogràfiques i els enllaços electrònics. Finalment s'insereixen les capçaleres, els panells de navegació i els





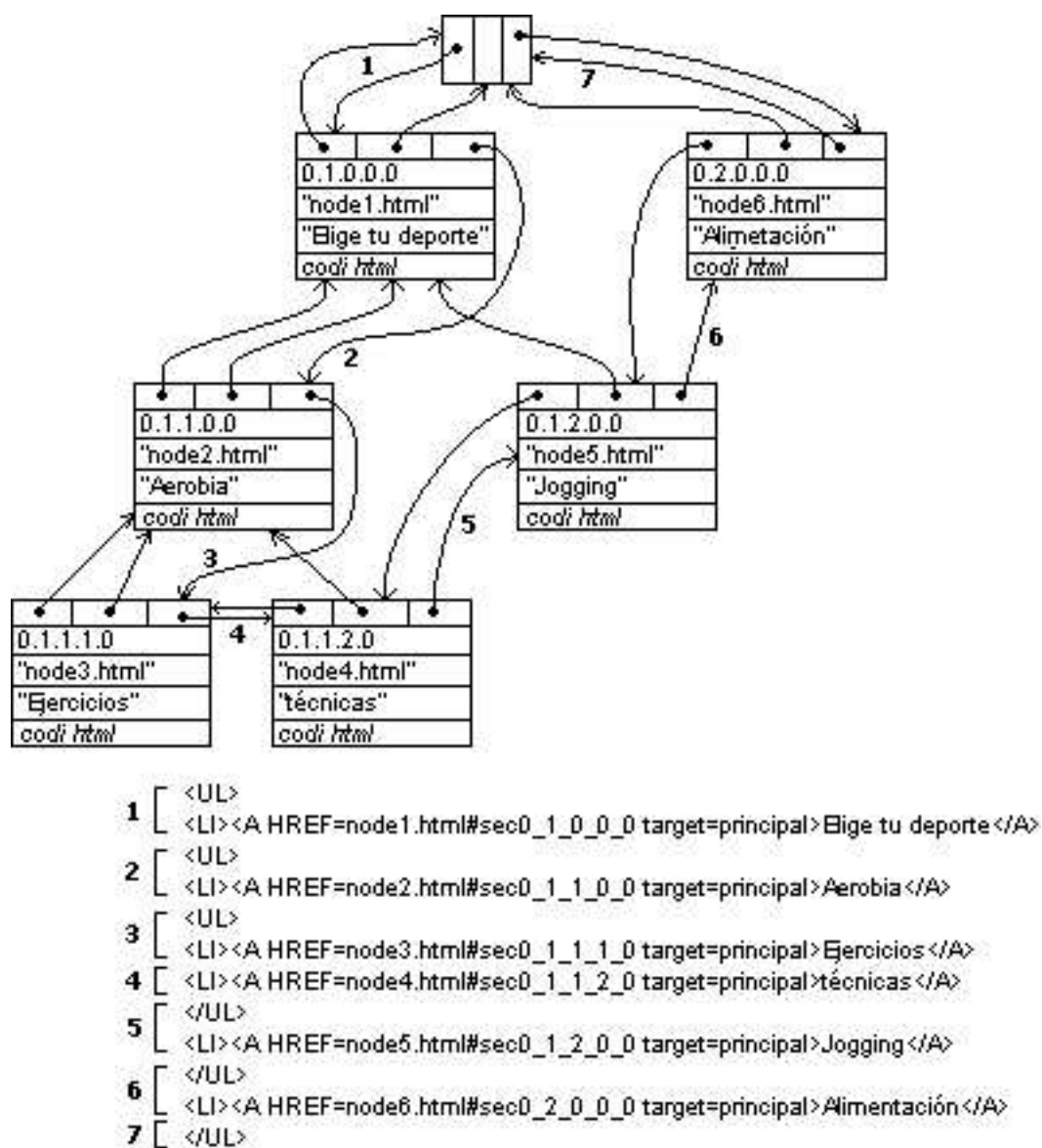


Figura 4.10: void aseccions::CrearTaulaContinguts()

peus a cada node i es procedeix a la materialització de cada un.

Quan el traductor localitza una macro `textbackslashpart`, `\chapter`, `\section`, `\subsection` o `\subsubsection` s'insereix l'apartat anterior a l'arbre mitjançant el mètode `document.novaSeccio(int tipus, char* titol, char* cos)` on `tipus` és un enumerat definit de la següent forma...

```
enum tipus_seccio {part=0, capitol=1, seccio=2, subseccio=3,
                  subsubseccio=4};
```

4.12 Subconjunt paragraph i subparagraph

Les marques `\paragraph` i `\subparagraph` defineixen un paràgraf i subparàgraf amb els seus títols respectivament. Tot i tractar-se d'un element estructural d'un document, per al traductor no es considera un apartat estructurable. La traducció de la marca `\paragraph` es realitza mitjançant la etiqueta `<P>` en HTML de la següent manera:

L^AT_EX

```
\paragraph{t\{'\i}tol}Spyglass Hill and Del Monte are the properties
courses in Monteray.
```

HTML

```
<p><b>t&#237;tol</b>&nbsp;&nbsp; Spyglass Hill and Del Monte are
the properties courses in Monteray.</p>
```

Representació

títol Spyglass Hill and Del Monte are the properties courses in Monteray.

La marca `\subparagraph` es realitza de la mateixa manera, però inclou un espaiat entre l'inici del paràgraf i el títol.

L^AT_EX

```
\subparagraph{t\{'\i}tol}Spyglass Hill and Del Monte are the properties
courses in Monteray.
```

HTML

```
<p>&nbsp;&nbsp; <b>t&#237;tol</b>&nbsp;&nbsp; Spyglass Hill and Del
Monte are the properties courses in Monteray.</p>
```

Representació

títol Spyglass Hill and Del Monte are the properties courses in Monteray.

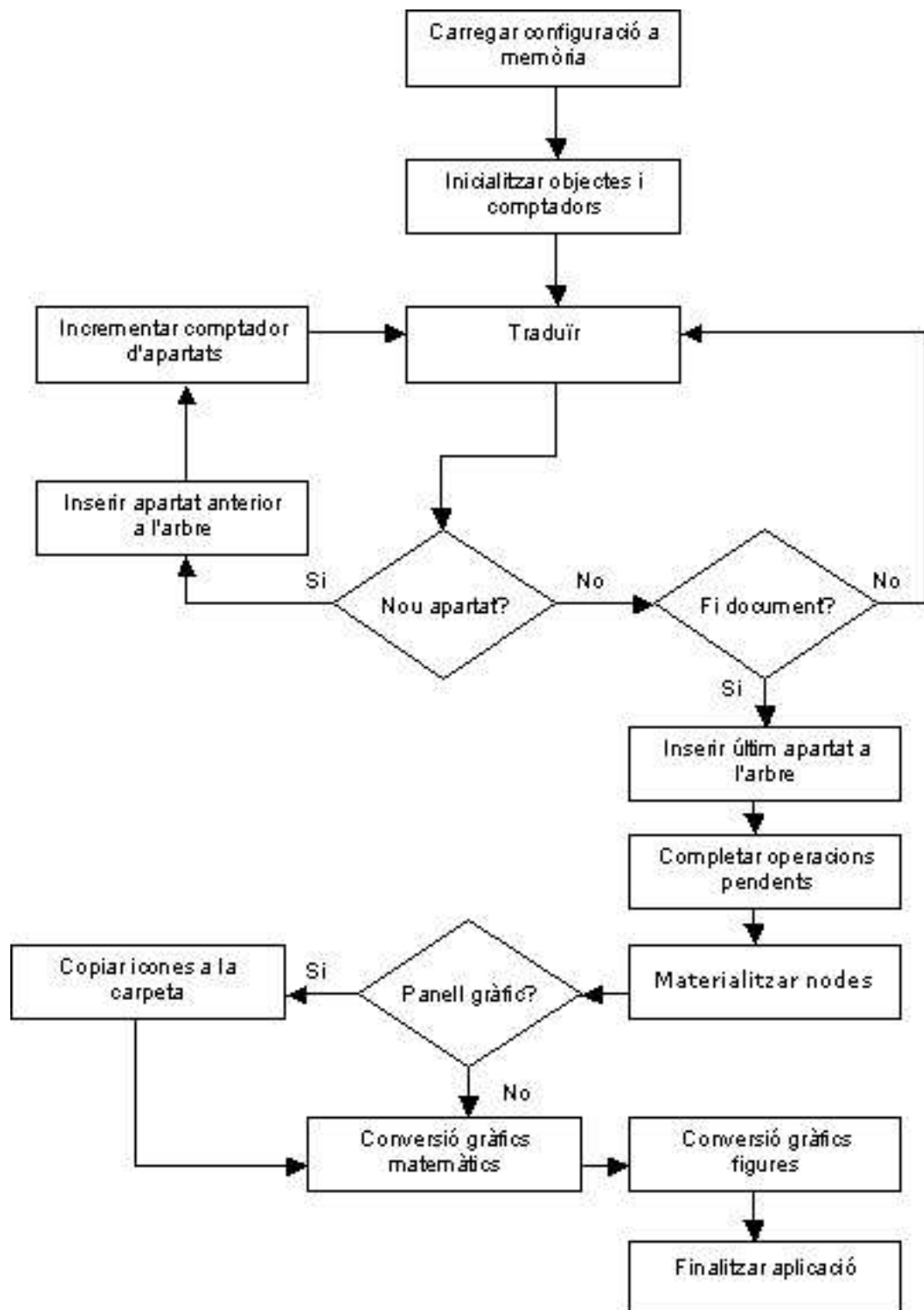


Figura 4.11: Diagrama general

4.13 Entorn `itemize` i `enumerate`

En \LaTeX es poden compondre enumeracions o llistes i disposar dels elements formant un paràgraf apart, mitjançant els entorns `itemize` i `enumerate`. Dins d'aquests entorns, cada element de l'enumeració ha d'anar precedit de la macro `item`.

Els entorns `itemize` i `enumerate` tenen el mateix comportament que les llistes i els enumerats corresponents al llenguatge HTML, i la seva traducció és pràcticament trivial, ja que simplement s'ha de reemplaçar unes marques per unes altres.

En el cas que afecta a l'entorn `itemize`, la composició de les llistes amb els seus corresponents nivells i subnivells es realitza anidant entorns de tipus `itemize`. Aquesta metodologia s'aplica tant en \LaTeX com en HTML i per tant, el problema es redueix simplement en crear nous nivells, inserir *items* i tancar els nivells oberts. El procés de traducció s'explica en els següents punts:

1. Quan el traductor troba l'inici d'un entorn `itemize` aleshores insereix en el codi HTML la marca d'obertura `` i es porten a terme els punts 2 i 3.
2. Mentre no es localitzi una macro `item` el traductor segueix traduïnt, en cas contrari, si es troba la macro el traductor insereix en el codi HTML la marca `` seguit del text traduït corresponent a l'ítem en qüestió.
3. Quan el traductor troba el final de l'entorn `itemize`, insereix en el codi HTML la marca de tancament `` i finalitza la llista o subnivell d'una llista.

En el cas de l'entorn `enumerate`, els elements de l'enumeració s'identifiquen mitjançant nombres aràbics, lletres minúscules incloses entre parèntesis, nombres romans en minúscula o bé mitjançant lletres en majúscula, segons es tracti d'enumeracions de primer nivell, segon, tercer o quart, respectivament. \LaTeX admet fins a un màxim de quatre nivells d'enumeracions, els quals no han de ser necessàriament tots entorns `itemize` o tots entorns `enumerate`.

HTML permet definir el tipus d'enumeració dels nivell o subnivells d'una llista. Per a fer possible això, s'ha d'indicar el tipus d'enumeració a les marques d'obertura de cada nivell, que en aquest cas tenen el format `<OL TYPE=*>`, on l'asterisc correspon al tipus d'enumeració i ha de prendre el valor **1** en el cas de l'enumeració aràbica per al nivell 1, el valor **a** en el cas de l'enumeració amb lletres minúscules entre parèntesi per al nivell 2, el valor **i** en el cas de l'enumeració en nombres romans en minúscula per al nivell 3 i el valor **A** en el cas de l'enumeració amb lletres majúscules per al nivell 4.

En aquest sentit, les llistes i enumeracions en el llenguatge **HTML** tenen exactament el mateix comportament, i quan es superen aquests quatre nivells tant **L^AT_EX** com un navegador o browser identifica els elements de la llista o enumeració de la mateixa forma que s'identifiquen els elements del nivell quatre.

El procés de traducció dels entorns **enumerate** s'explica en els següents punts:

1. Quan el traductor troba l'inici d'un entorn **enumerate**, aleshores insereix en el codi **HTML** la marca d'obertura `<OL TYPE=*>`, on el caràcter que ha d'anar en la posició de l'asterisc depèn del valor de la variable *nivell*. S'incrementa en una unitat el valor de la variable *nivell*. Es porten a terme els punts 2 i 3.
2. Mentre no es localitzi una macro **item** el traductor segueix traduïnt, en cas contrari, si es troba la macro el traductor insereix en el codi **HTML** la marca `` seguit del text traduït corresponent a l'ítem en qüestió.
3. Quan el traductor troba el final de l'entorn **enumerate**, insereix en el codi **HTML** la marca de tancament ``, decrementa en una unitat el valor de la variable *nivell* i finalitza l'enumeració o subnivell d'una enumeració.

A la figura 4.12 es pot observar amb més detall el diagrama corresponent al procés de traducció d'entorns **enumerate**.

4.14 Entorn *tabbing*

L'entorn *tabbing* ofereix la possibilitat d'aliniat text en columnes. En aquest entorn, l'alineació del text en columnes s'especifica fixant posicions de tabulació (és a dir, fixant l'espaiat automàtic del tabulador) i fent un salt fins a la pròxima posició de tabulació (és a dir, aturant el carro a la pròxima columna del tabulador).

Hi ha dos maneres d'especificar les posicions de tabulació en un entorn *tabbing*: en una línia del text que s'ha de compondre en columnes o bé, de manera explícita, fixant l'espaiat fins a la pròxima posició de tabulació. En qualsevol cas, dins de l'entorn *tabbing*, la macro `\=` fixa una posició de tabulació i la macro `\>` indica un salt fins a la pròxima posició de tabulació. Les línies de la taula han d'estar separades per macros de final de línia `\\`. D'aquesta manera, el següent exemple en **L^AT_EX**...

```
\begin{tabbing}
\textbf{función} \=fib(n: entero): entero \\
\> \textbf{si} \= n=0 \textbf{entonces} \\
\> \> retorna(0) \\
\> \textbf{sino} \\
```

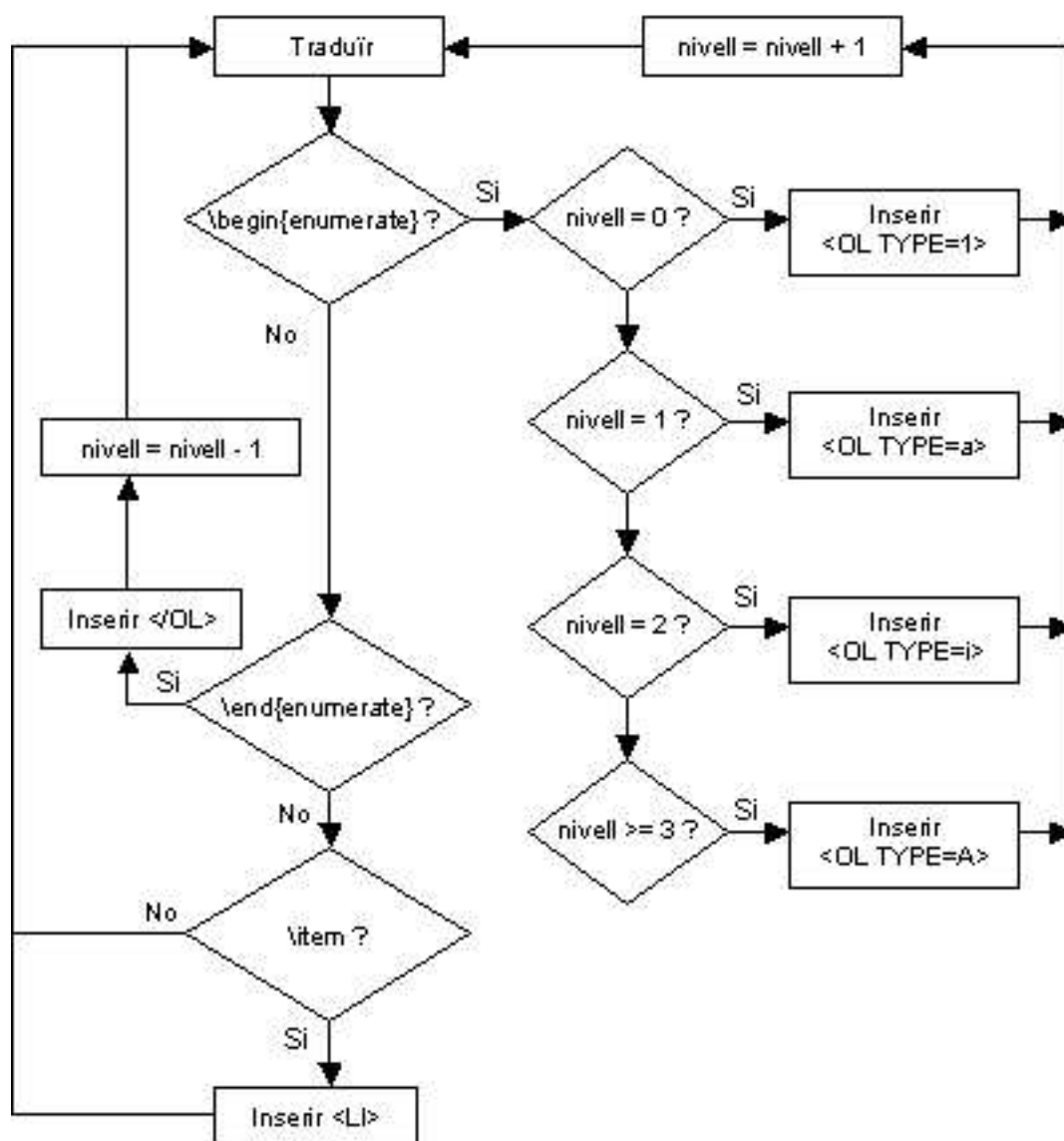



Figura 4.12: Diagrama de traducció d'entorns `enumerate`

```

\> \> \textbf{si} \= n=1 \textbf{entonces} \\
\> \> \> retorna(1)\\
\> \> \textbf{sino} \\
\> \> \> retorna(fib(n-1)+fib(n-2)) \\
\> \> \textbf{fsi} \\
\> \textbf{fsi} \\
\textbf{ffunción}\\
\end{tabbing}

```

...genera el següent resultat...

```

funci fib(n: entero): entero
    si n=0 entonces
        retorna(0)
    sino
        si n=1 entonces
            retorna(1)
        sino
            retorna(fib(n-1)+fib(n-2))
        fsi
    fsi
funci

```

Un cop vist això, anem a veure com el traductor transforma aquesta classe de textos a HTML. El traductor contempla un màxim de 20 tabulacions definides per la macro `\=` en un entorn `tabbing`, d'aquesta manera un vector d'enters de 20 posicions s'encarrega d'emmagatzemar l'espaiat corresponent a cada tabulació. En aquest sentit, l'espai de cada tabulació és l'espai en caràcters que hi ha entre la marca `\=` i l'origen de la línia.

Quan finalitza una línia (és a dir, quan el traductor detecta la macro `\\` dins de l'entorn `tabbing`) es torna a inicialitzar a 0 la variable encarregada del compteig del nombre de caràcters. Així doncs, la traducció a HTML consisteix en reemplaçar les macros `\>` pel nombre d'espais corresponent i aplicar un trenca-ment de línia `
` quan es detecta la macro `\\`. En finalitzar l'entorn `tabbing` es restauren els valors de les variables implicades al seu valor inicial.

Els textos tabulats definits per l'entorn `tabbing` tenen la característica de tenir un tipus d'alineació diferent de la resta d'elements que componen un text. No té sentit centrar o alinear a la dreta el text d'un entorn `tabbing`, ja que sinó les tabulacions no funcionen correctament. El que fa el sistema `LATEX` en aquests casos, és centrar el text, però mantenint les tabulacions i l'origen d'aquestes. Per simular aquest efecte i, evitar així el problema anterior, el traductor insereix la traducció dels entorns `tabbing` dins d'una taula, d'aquesta manera quan s'aplica un entorn `center` o `flushright` a un entorn `tabbing`, en realitat l'alineació té efecte sobre la taula i no en el text i per tant, el problema es resol satisfactòriament.

LaTeX		HTML
<pre> \begin{tabbing} vo\=id\= prova()\ \> {} \> \> a = a + 1;)\ \end{tabbing> </pre>		
<p>↓</p> <p>1 vo\=id\= prova()\</p> <pre> espais=2 tab=0 tab_compt=0 tab_espais = {0,0,0,...,0} </pre>		<p>↓</p> <p>vo</p>
<p>↓</p> <p>2 vo\=id\= prova()\</p> <pre> espais=0 tab=1 tab_compt=0 tab_espais = {2,0,0,...,0} </pre>		<p>vo</p>
<p>↓</p> <p>3 vo\=id\= prova()\</p> <pre> espais=2 tab=1 tab_compt=0 tab_espais = {2,0,0,...,0} </pre>		<p>↓</p> <p>voidid</p>
<p>↓</p> <p>4 vo\=id\= prova()\</p> <pre> espais=0 tab=2 tab_compt=0 tab_espais = {2,2,0,...,0} </pre>		<p>void</p>
<p>↓</p> <p>5 vo\=id\=prova()\</p> <pre> espais=7 tab=2 tab_compt=0 tab_espais = {2,2,0,...,0} </pre>		<p>↓</p> <p>voidprova()</p>
<p>↓</p> <p>6 vo\=id\= prova()\</p> <pre> espais=0 tab=2 tab_compt=0 tab_espais = {2,2,0,...,0} </pre>		<p>↓</p> <p>void prova()
</p>
<p>↓</p> <p>7 \> {}</p> <pre> espais=0 tab=2 tab_compt=1 tab_espais = {2,2,0,...,0} </pre>		<p>void prova()
</p> <p> ←</p>

8	<div> <div>↓</div> <div>\> { \ \</div> <div> <div>espais=1 tab=2 tab_compt=1</div> <div>tab_espais = {2,2,0,...,0}</div> </div> </div>	<div> <div>void prova()
</div> <div>{ ←</div> </div>
9	<div> <div>↓</div> <div>\> { \ \</div> <div> <div>espais=0 tab=2 tab_compt=0</div> <div>tab_espais = {2,2,0,...,0}</div> </div> </div>	<div> <div>void prova()
</div> <div>{
 ←</div> </div>
10	<div> <div>↓</div> <div>\> \> a = a + 1; \ \</div> <div> <div>espais=0 tab=2 tab_compt=1</div> <div>tab_espais = {2,2,0,...,0}</div> </div> </div>	<div> <div>void prova()
</div> <div>{
</div> <div>←</div> </div>
11	<div> <div>↓</div> <div>\> \> a = a + 1; \ \</div> <div> <div>espais=0 tab=2 tab_compt=2</div> <div>tab_espais = {2,2,0,...,0}</div> </div> </div>	<div> <div>void prova()
</div> <div>{
</div> <div>←</div> </div>
12	<div> <div>↓</div> <div>\> \> a = a + 1; \ \</div> <div> <div>espais=11 tab=2 tab_compt=2</div> <div>tab_espais = {2,2,0,...,0}</div> </div> </div>	<div> <div>void prova()
</div> <div>{
</div> <div>a = a + 1; ←</div> </div>
13	<div> <div>↓</div> <div>\> \> a = a + 1; \ \</div> <div> <div>espais=0 tab=2 tab_compt=0</div> <div>tab_espais = {2,2,0,...,0}</div> </div> </div>	<div> <div>void prova()
</div> <div>{
</div> <div>a = a + 1; ↓
</div> </div>

4.15 Entorn figure i table

Els textos científics solen incloure moltes figures i taules de dades, la composició de les quals comporta el problema de que, a diferència dels paràgrafs de text, les figures i les taules normalment no es poden dividir. L^AT_EX soluciona aquest problema començant una nova pàgina cada vegada que una figura o una taula es massa gran per compondre-la dins de la pàgina. En aquests casos, es diu que les figures i les taules *floten* en el document compost fins que el L^AT_EX troba espai lliure suficient per incloure-les. Tot i ser una solució molt simple, té l'inconvenient de que acostuma a deixar moltes pàgines mig buides, cosa que no agrada ja sigui per raons econòmiques o per raons d'estètica i claredat d'exposició del text.

Aquesta problemàtica que tenen les figures i les taules desapareix alhora de traduir un document L^AT_EX a HTML, ja que el concepte de pàgina no té sentit en una traducció estructurada per apartats, on cada node contempla el contingut total de l'apartat que té assignat.

Els entorns **figure** i **table** són molt semblants, ja que tenen finalitats pràcticament idèntiques. Una figura és una il·lustració, és a dir, un dibuix, una làmina, una pintura, una fotografia, un mapa, etc., que il·lustra el text d'un document. L'entorn **figure** té el següent format...

```
\begin{figure}
figura
\end{figure}
```

Dins de l'entorn, *figura* pot ser qualsevol paràgraf o paràgrafs de text, entorns matemàtics, altres entorns de text, dibuixos i il·lustracions produïts per altres eines gràfiques. Les figures solen incloure una llegenda que descriu el contingut, a més a més d'una etiqueta que permet fer referències creuades dins del text del document, cosa que s'aconsegueix mitjançant les macros estàndars de referències creuades **caption** i **label**, respectivament.

La llegenda d'una figura es compona sempre per sota de la figura i, en aquest traductor proposat, la llegenda definida per la macro **caption** ha d'estar sempre fora de qualsevol entorn d'alineació de text, ja que la llegenda sempre es representa amb alineació centrada. Per a que l'enumeració de les referències creuades sigui correcta, és important incloure l'etiqueta després de la llegenda de la figura, o bé escriure-la dins de la llegenda, però mai s'ha de posar abans.

```
\begin{figure}
figura
\caption{llegenda}
\label{etiqueta}
\end{figure}
```

Les llegendes de les figures surten a l'índex de figures del document L^AT_EX per mitjà de la macro **listoffigures**. Per fer possible el mecanisme de referències

creuades en les figures, s'assigna a l'inici de l'entorn una etiqueta a la figura i si el traductor troba la macro `label` dins de l'entorn, aleshores s'associa l'etiqueta amb l'etiqueta definida per `label` per mitjà de la pila d'entorns. El procés de traducció dels entorns `figure` s'explica a continuació:

1. Quan el traductor troba l'inici d'un entorn `figure` aleshores crea una etiqueta amb el format `F*`, on l'asterisc representa el número de figura en curs, i insereix en el codi HTML el marcatge d'obertura `<P>`. Afegeix a la pila d'entorns l'etiqueta de la figura i el nom de fitxer associat a l'apartat on es troba la figura, i incrementa en una unitat el comptador de figures. Es porten a terme els punts 2 i 3.
2. Mentre no es localitzi la macro `caption` el traductor segueix traduint, en cas contrari, si es troba la macro el traductor insereix en el codi HTML el següent fragment: `<CENTER>Figura *: llegend</CENTER>`, on l'asterisc representa el número de figura en curs i `llegend` representa la llegenda en format textual.
3. Quan el traductor troba la macro de final de l'entorn `figure`, insereix en el codi HTML la marca de tancament de paràgraf `</P>`.

La figura 4.13 mostra com funciona el mecanisme de referències creuades amb les figures.

La inclusió de fitxers gràfics s'aconsegueix mitjançant els fitxers gràfics en format PostScript(PS) o PostScript encapsulat (EPS), un format difós en prcticament tots els ordinadors. De fet, quasi totes les eines informàtiques de dibuix i elaboració de gràfics permeten generar fitxers gràfics en format PS o EPS. Aquests fitxers es poden incloure en un document \LaTeX mitjançant l'activació de paquets de definicions, entre els quals cal destacar el paquet `graphics` i el paquet `epsfig`. Aquests dos paquets de definicions admeten una opció que especifica el traductor gràfic de pantalla i d'impressora, com ara `dvips` i `xdvi`, per a les implementacions dels sistemes \TeX i \LaTeX .

Amb els paquets `graphics` i `epsfig` les macros `\includegraphics{fitxer}` i `\epsfig{file=fitxer}` inclouen el contingut del `fitxer` en el document \LaTeX compost, on `fitxer` és el nom del fitxer gràfic en PS o EPS. Cal dir que el traductor només contempla el paràmetre `file` per a la macro `epsfig`.

És important indicar que el traductor no exigeix declarar els paquets `graphics` i `epsfig`, ni cap altre tipus de paquets, a l'inici del document \LaTeX , ja que directament es tradueixen les macros sense considerar si són vàlides o no. En aquests casos, se sobreentèn que l'usuari ha redactat correctament el document \LaTeX .

LaTeX	HTML
<pre> \begin{figure} \begin{center} \includegraphics{foto.ps} \end{center} \caption{La meva foto personal} \label{foto} \end{figure} ... La figura \ref{foto} és la meva foto. </pre>	<pre> <P> <CENTER> </CENTER> <CENTER> Figura 3: La meva foto personal </CENTER></P> ... <P>La figura 3 &#233;s la meva foto.</P> </pre>

Representació en un navegador o browser

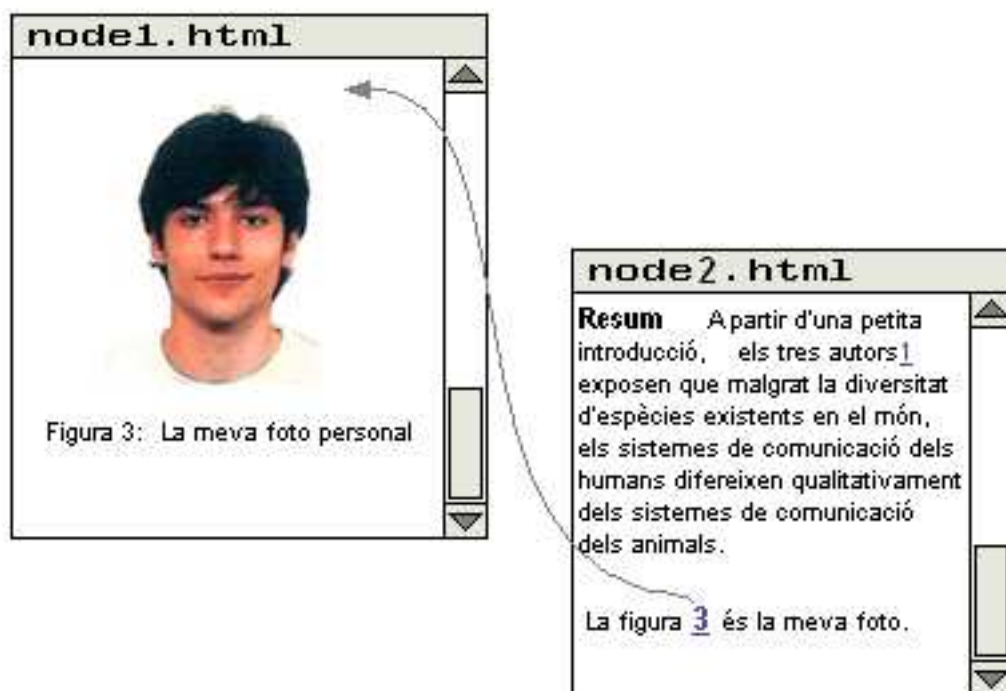


Figura 4.13: Exemple de referències creuades i figures

Una taula és una llista sistemàtica de dades que presenta de manera clara i concisa una quantitat relativament gran de dades, la descripció de les quals necessitaria d'una altra forma molts paràgrafs de text. Les taules són principalment numèriques, tot i que una taula de paràgrafs de text pot contribuir a simplificar l'exposició d'un text complex.

Les taules que s'usen amb més freqüència en els textos *acadèmic-científics* són les taules estadístiques i les taules de funcions. Una taula estadística presenta les variacions d'una o més categories de dades (*variables dependents*) segons altres categories de dades (*variables independents*).

En \LaTeX les taules es componen mitjançant l'entorn `table` de la següent forma:

```
\begin{table}
taula
\end{table}
```

Dins d'aquest entorn, *taula* pot ser qualsevol paràgraf o paràgrafs de text, entorns matemàtics, altres entorns de text, etc., tot i que normalment s'hi sol incloure un entorn `tabular` com a contingut. De la mateixa manera que les figures, les taules també acostumen a incloure una llegenda que descriu el seu contingut, a més a més d'una etiqueta que permet fer referències creuades dins del text del document, cosa que s'aconsegueix mitjançant les macros estàndars de referències creuades `caption` i `label`, respectivament. Per a que l'enumeració de les referències creuades sigui correcta, s'ha d'incloure l'etiqueta després de la llegenda de la taula, o bé posar-la dins de la llegenda, però mai abans. A diferència de les figures, la llegenda de les quals es compona a sota, la llegenda d'una taula es compona sempre a sobre de la taula.

4.16 Macros `tableofcontents`, `listoffigures` i `listoftables`

La macro `tableofcontents` especifica la composició de l'índex d'un document de la classe `article`, `report` o `book`, tot i que el traductor no contempla el tipus de classe del document que tradueix. \LaTeX compona la taula de continguts partint dels títols dels apartats i dels números de pàgina corresponents, els quals són un subproducte de la última composició del document, i el \LaTeX els obté del fitxer auxiliar amb extensió `.toc` (table of contents). Això vol dir que \LaTeX processa el document almenys dos vegades per obtenir un índex correcte.

La profunditat o número de nivells de títols de capítol i apartat que surten a l'índex ve donada per un paràmetre intern de \LaTeX . En aquest sentit, en el compilador \LaTeX , el nombre màxim de nivells representables a la taula de continguts és, per defecte, igual a 3 per a totes les classes estàndars de documents. D'aquesta manera, a l'índex d'un document de classe `book` o `report` sortiran

els títols de *capítol*, *secció* i *subsecció*, mentre que els títols de *secció*, *subsecció* i *subsubsecció* sortiran a l'índex d'un document de la classe **article**.

El traductor segueix una política general de traducció, evitant traduir en funció d'una classe o una altra de document, simplement es tradueix seguint els criteris de traducció establerts. Aleshores el format de la taula de continguts, i per tant, l'estructuració del document HTML final està definit segons la configuració que l'usuari ha establert al fitxer de configuració (documentat amb més detall a l'apartat A.4).

El traductor contempla una profunditat per defecte de 5 nivells, això vol dir que si un document conté, per exemple, **parts**, **capítols**, **seccions**, **subseccions** i **subsubseccions** aleshores es generaran nodes HTML per cada apartat que pertany a qualsevol dels apartats mencionats, i conseqüentment, la taula de continguts tindrà 5 nivells. S'observa que la relació entre la taula de continguts i l'estructuració general dels documents HTML generats és clara: la taula de continguts representa la forma en què està estructurat el document, i per tant, el document està fragmentat en tants nodes com número d'apartats hi ha a la taula de continguts.

Com ja s'ha mencionat, la profunditat de la taula de continguts que genera el traductor és 5 per defecte, i aquesta es pot modificar ajustant les propietats del fitxer de configuració. L'usuari pot decidir si vol que un tipus d'apartat aparegui a la taula de continguts o no.

En aquest sentit, si l'usuari no vol que apareguin els capítols, el traductor generarà una taula de continguts d'una profunditat màxima de 4 nivells, ja que els capítols no es representaran. Això no vol dir que els continguts dels capítols desapareguin o es menysprein en el document HTML final, sinó que el traductor es limitarà a inserir els continguts dels capítols al final del node de l'apartat de nivell superior, o a l'inici del node de l'apartat de nivell inferior si no es possible el cas anterior. Això generarà una taula de continguts més petita i el traductor crearà menys nodes HTML.

Per fer que el traductor adopti aquest comportament només s'ha d'evitar inserir a l'arbre d'apartats aquells apartats que no volem que formin part de la taula de continguts. D'aquesta manera podem fer, per exemple, que a la taula de continguts només apareguin *parts* i *seccions*, indicant que no volem que els *capítols*, *subseccions* i *subsubseccions* en formin part. La figura 4.14 en mostra un clar exemple.

La macro `listoffigures` especifica la composició de l'índex de figures del document. \LaTeX ho compona partint dels títols de l'enumeració de les figures, de les llegendes i dels números de pàgina corresponents, que són un subproducte de l'última composició del document. El traductor quan localitza la macro `listoffigures` ho tradueix inserint la cadena `%F` al codi HTML resultant. Aque-

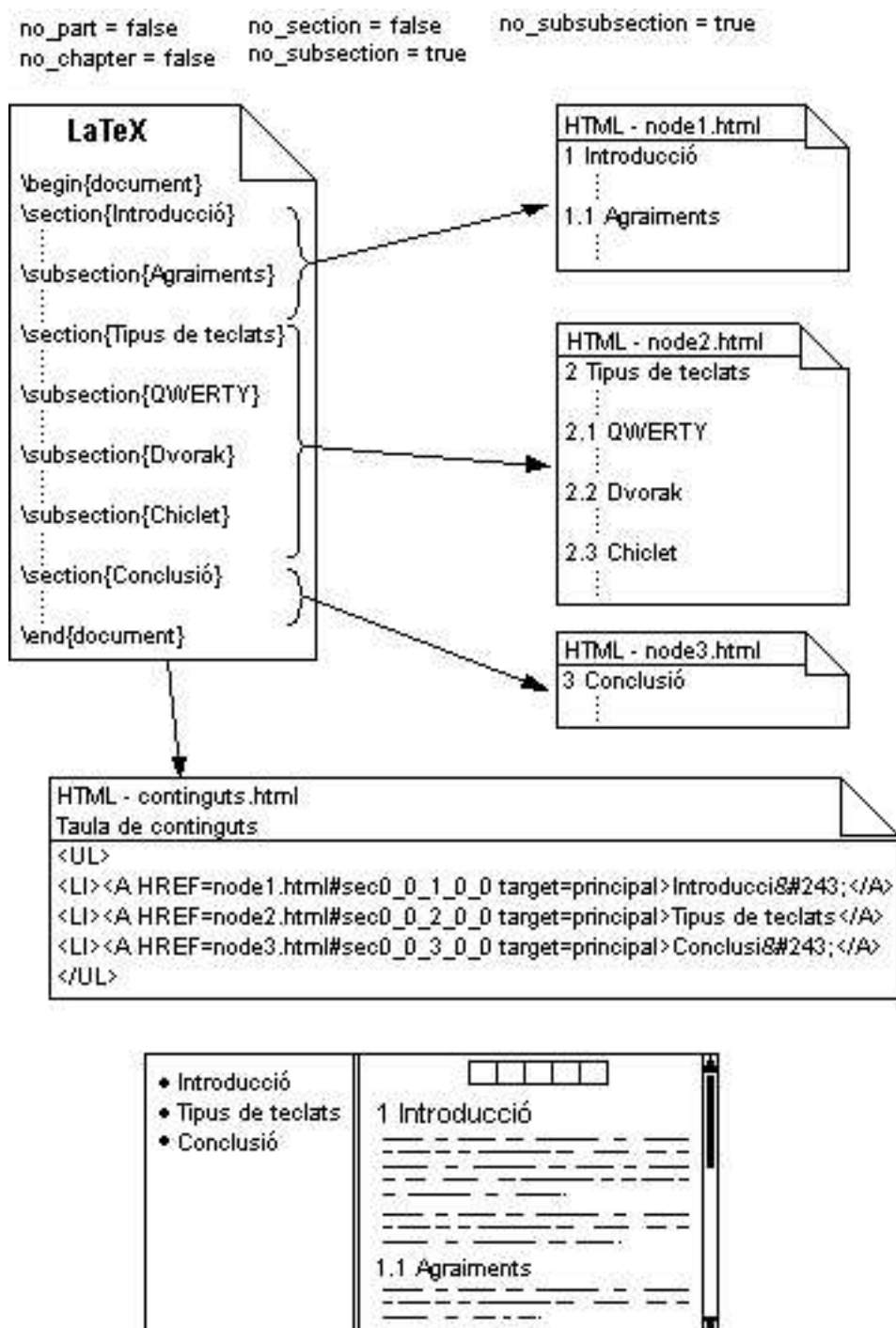


Figura 4.14: Exemple d'una taula de continguts configurable

sta cadena és de caràcter reservat, això vol dir que aquesta cadena és informativa i per tant, indica que hi ha una operació pendent: afegir la llista de figures al codi **HTML** un cop hagi finalitzat la traducció en curs.

La macro **listoftables** especifica la composició de l'índex de taules del document. **L^AT_EX** ho compona partint dels títols de l'enumeració de les taules, de les llegendes i dels números de pàgina corresponents, que són un subproducte de l'última composició del document. Quan el traductor localitza la macro **listoftables** realitza la mateixa operació que amb el cas de les llistes de figures, però en aquest cas la cadena de caràcters que insereix al codi **HTML** és **%T**. En el cas de la macro **tableofcontents**, la mecànica és idèntica, però la cadena que s'insereix és **%C**.

Així doncs, les macros **listoftables**, **listoffigures** i **tableofcontents** són operacions pendents que s'han de completar un cop finalitzada la traducció del document. Aquestes operacions les porta a terme la classe **aseccions**, i es completen substituint les cadenes **%C**, **%T** i **%F** pel codi **HTML** generat per a cada un.

Capítol 5

Conversió gràfica de fragments matemàtics i figures

La traducció d'un document \LaTeX consisteix tant en la traducció del propi text com de la traducció de possibles fragments matemàtics i figures incloses en aquest. No hi ha uns criteris clars per traduir fragments matemàtics de \LaTeX a HTML, ja que HTML no disposa de suficients mecanismes per representar expressions matemàtiques mínimament complexes.

Una alternativa proposada inicialment ha estat l'ús combinat de taules, línies i operadors matemàtics disponibles en el llenguatge HTML, però amb això només s'aconsegueix representar un conjunt concret d'expressions, a més a més, aquesta possibilitat inclou una problemàtica afegida, i és que aleshores s'ha de saber identificar si una expressió matemàtica en \LaTeX és representable o no, amb aquests mitjans.

L'alternativa més fiable és la conversió dels fragments matemàtics a un format gràfic que pugui ser visualitzat a través d'un navegador o browser convencional. El tractament de les figures és similar al dels fragments matemàtics, però en aquest cas només fa falta la conversió directa d'un format gràfic a un altre. En els següents apartats s'explica detalladament com es porten a terme aquests processos de conversió durant la traducció.

5.1 Conversió de fragments matemàtics

Tal i com s'explica en l'apartat anterior, la traducció dels fragments matemàtics es fa per mitjà d'una conversió a un format gràfic que pugui ser visualitzat a través d'un navegador web o browser convencional. Per a fer possible aque-

sta conversió s'utilitza un compilador de \LaTeX en qualsevol de les seves distribucions, sempre i quan faci possible la traducció de documents al format **PostScript**.

En aquest cas, interessa únicament traduir el fragment o fragments matemàtics al format **PostScript** i posteriorment la seva adient conversió del fitxer **PostScript** al format **GIF**, mitjançant les eines de conversió gràfiques d'un interpret de llenguatge **PostScript** com és **GhostScript**. Aquests passos es porten a terme en un conjunt de fases que es descriuen a continuació:

1. Es captura el fragment matemàtic inclòs entre les marques $\$$ o $\$\$$
2. El fragment s'escriu en un fitxer de text seguint el format estàndar d'un document \LaTeX .
3. Es compila el fitxer amb un compilador \LaTeX i s'obté un fitxer **PostScript** com a resultat de l'operació.
4. S'utilitza **GhostScript** per a fer possible la conversió del fitxer en format **PostScript** a un altre en format **GIF**, i s'eliminen la resta de fitxers utilitzats durant tot el procés.
5. Es guarda el fitxer **GIF** resultant al directori destí.

Els passos 3 i 4 es realitzen mitjançant els scripts *latex2gif* i *ps2gif* implementats per a aquests casos. L'ús d'scripts permet a l'usuari realitzar modificacions que afecten a les conversions gràfiques sense haver d'interferir en el codi font del programa. El codi de l'script *latex2gif* per a la distribució de \LaTeX *vtex* és el que es mostra a continuació:

```
#!/bin/sh
vlatex -q -n1 @latex $1.tex > /dev/null
./ps2gif $1.ps $2.gif 140
rm $1.tex
rm $1.aux
rm $1.log
rm $1.ps
```

El codi de l'script *latex2gif* per a la distribució de \LaTeX *web2C* estàndar és el següent:

```
#!/bin/sh
latex -interaction=batchmode $1.tex > /dev/null
dvips $1.dvi -o $1.ps -q > /dev/null
./ps2gif $1.ps $2.gif 140
rm $1.tex
rm $1.aux
rm $1.dvi
rm $1.log
rm $1.ps
```

La gestió, tant de conversions gràfiques de fragments matemàtics, com de figures, la porta a terme un objecte de la classe `conversorGrafic`. Aquesta classe s'encarrega del manteniment de dos cues que emmagatzemen fragments matemàtics i figures respectivament, a més a més, d'executar els scripts *latex2gif* i *ps2gif*. Les operacions principals per al tractament dels fragments matemàtics que encapçula aquesta classe es detallen a continuació:

`char* conversorGrafic::inserirFragmentMatematic(char* fragment)` Afegeix el fragment matemàtic *fragment* a una cua. Retorna el nom del fitxer gràfic que generarà en format GIF.

`void conversorGrafic::iniciarConversioMatematica()` Inicia la conversió de tots els fragments matemàtics inserits a la cua.

5.2 Conversió de figures

La conversió de figures consisteix en la conversió de fitxers en format PS o EPS al format gràfic GIF. Aquest procés es realitza en un únic pas mitjançant una crida a l'script *ps2gif* per a cada figura que el traductor ha de convertir. És important remarcar que el traductor realitza les conversions, tant de figures com de fragments matemàtics, al final de la traducció, donat que la traducció bàsica del document és prioritària.

El codi de l'script *ps2gif* és genèric i depèn en aquest cas de l'interpret de llenguatge PostScript, GhostScript. El codi és el següent:

```
#!/bin/sh
#Requereix tenir Ghostscript i les utilitats pbmplus instal·lades
echo "["$2"]>&2
gs -sDEVICE=ppmraw -sOutputFile=- -sNOPAUSE -r$3 -q $1 -c
showpage -c quit | pnmcrop -quiet | pnmmargin -white 10 | ppmquant
256 -quiet | ppmtogif -quiet >$2
```

En el cas de la conversió de les figures, les operacions principals que encapçula la classe `conversorGrafic` es detallen a continuació:

`void conversorGrafic::void inserirImatge(char* img)` Afegeix el nom de fitxer *img* a una cua. No retorna cap valor ja que en aquest cas el nom de fitxer que generarà serà el mateix però en extensió GIF.

`void conversorGrafic::iniciarConversioImatges()` Inicia la conversió de tots els fitxer inserits a la cua.

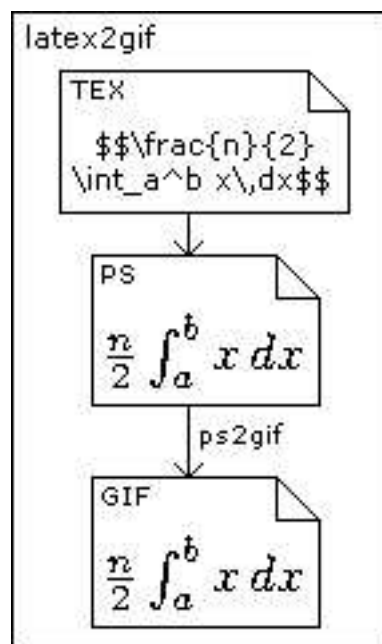


Figura 5.1: Conversió gràfica de fragments matemàtics

Capítol 6

Presentació i navegació

El format dels documents traduïts ha estat, desde el primer moment, un factor d'important atenció. S'ha considerat interessant la possibilitat de que l'usuari tingui a l'abast un accés directe a qualsevol dels apartats que constitueixen els documents traduïts sense necessitat d'haver de recórrer tot, o part, del document fins arribar a la informació que li interessa.

6.1 Entorn amb marcs

Per satisfer aquest propòsit inicial s'ha optat per la solució que ofereixen els marcs (frames) del llenguatge HTML, amb els quals es possible fragmentar la presentació del document en dos parts independents. En una part es mostra la taula de continguts i en l'altra el contingut de la secció en curs seleccionada per l'usuari, d'aquesta manera és possible accedir amb un sol click a la secció concreta que es vol consultar.



Figura 6.1: Entorn amb marcs



Figura 6.2: Entorn amb marcs

La presentació del document traduït es mostra en el navegador web amb dos marcs, un a la dreta que conté l'apartat en curs i un a l'esquerra que conté la taula de continguts del document. Quan es fa click a un apartat concret del marc de l'esquerra s'ordena la càrrega i visualització del fitxer associat a aquell apartat en el marc de la dreta.

La taula de continguts del marc esquerra es presenta en el navegador seguint un esquema indexat, l'amplada del qual pot variar en funció de la longitud dels títols dels apartats i del nombre de nivells i subnivells que tingui. Els navegadors web solucionen el problema de la longitud dels textos, realitzant un trencament de línia quan el text supera al marge dret del marc, tanmateix això no succeeix amb l'estructura de les taules de continguts (llistes, numeracions, etc...), en aquest cas els navegadors amplien el tamany del marc mantenint l'àrea de presentació, això ho solucionen mitjançant barres de desplaçament.

Per evitar aquest molest problema, el traductor ofereix la possibilitat d'ajustar les propietats de la presentació per tal que aquesta pugui adaptar-se a les característiques del document. Aquestes propietats es poden corregir mitjançant una adequada configuració del fitxer de configuració del traductor. Entre totes les propietats configurables, és important destacar en aquest apartat, i de forma superficial, aquelles que afecten a la presentació del document HTML.

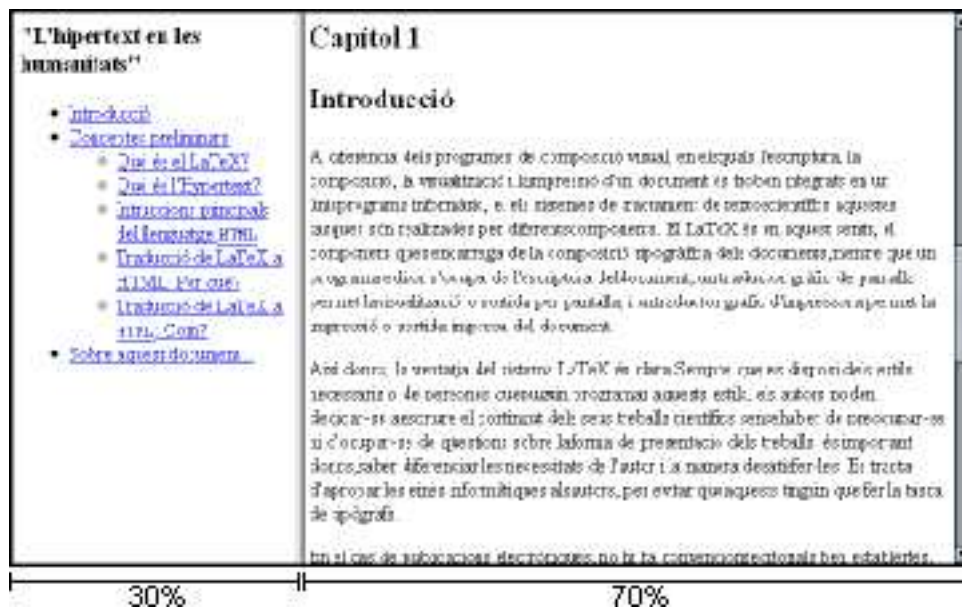


Figura 6.3: Marcs de longitud variable

Longitud dels marcs La longitud dels marcs és variable i aquesta es pot modificar indicant el percentatge del tamany del marc esquerra. Per defecte, el tamany del marc esquerra és el 30% de l'espai total representable en el navegador web, i per tant, el 70% restant correspon al marc de la dreta.

Ocultació de la barra de separació Per qüestions estètiques pot resultar interessant ocultar la barra central de separació dels marcs.

Presentació sense marc esquerra Es pot evitar crear el marc de l'esquerra i fer que la navegació a través del document es realitzi únicament per mitjà dels panells de navegació superior i inferior. El document es mostra en el 100% de l'amplada de la finestra.

Ús de fulles d'estil Es pot associar el document amb una fulla d'estil creada per l'usuari, d'aquesta forma el format del document el determina la fulla d'estil.

Color de fons Permet definir el color de fons dels documents HTML.

Color dels links Permet especificar el color dels links que encara no s'han visitat.

Links ja visitats Permet especificar el color dels links ja visitats.



Figura 6.4: Ocultar la separació dels marcs

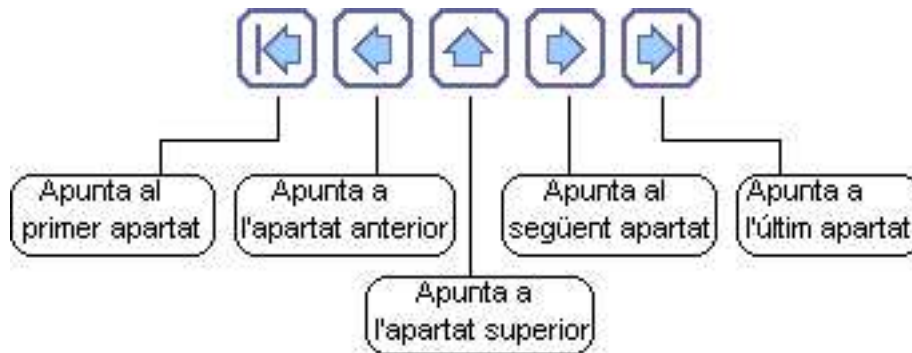


Figura 6.5: Panell de navegació

Links actius Permet especificar el color dels links en el moment que l'usuari fa click sobre d'ells.

6.2 Sistema de navegació

El sistema de navegació dissenyat consisteix en un panell de navegació format per 5 botons, cadascun condueix a un apartat determinat del document. El panell de navegació s'afegeix a l'inici i final de cada apartat i el seu format queda representat a la figura 6.5.

El **primer enllaç** apunta al primer apartat del document, és a dir, si el document està desglossat en capítols, el primer apartat serà el capítol 1. El **segon**



Figura 6.6: Diferents estils amb fulles d'estil diferents

[Primer](#) | [Anterior](#) | [Superior](#) | [Següent](#) | [Ultim](#)

Figura 6.7: Panell de navegació en mode text

enllaç apunta a l'anterior apartat, és a dir, si l'apartat en curs és la subsecció 2 de la secció 3 aleshores al *clicar* al damunt de l'enllaç accedirem a la subsecció 1 de la secció 3, i si tornem a clicar damunt de l'enllaç un altre cop, aleshores accedirem a l'arrel de la secció 3. El **tercer enllaç** apunta a l'apartat superior, o el que és el mateix, accedirem directament a l'arrel del nivell en curs. Si estem a la subsecció 2 de la secció 3, aleshores al *clicar* damunt de l'enllaç accedirem a l'arrel de la secció 3. El **quart enllaç** actua igual que el segon, però en aquest cas accedim al següent apartat. De la mateixa manera, el **cinquer** i últim enllaç actua igual que el primer, però en aquest cas accedim a l'últim apartat del document.

L'usuari pot decidir utilitzar un panell gràfic com el de la figura 6.5 o bé utilitzar un panell en mode text com el de la figura 6.7. Aquesta propietat es pot ajustar per mitjà del fitxer de configuració.

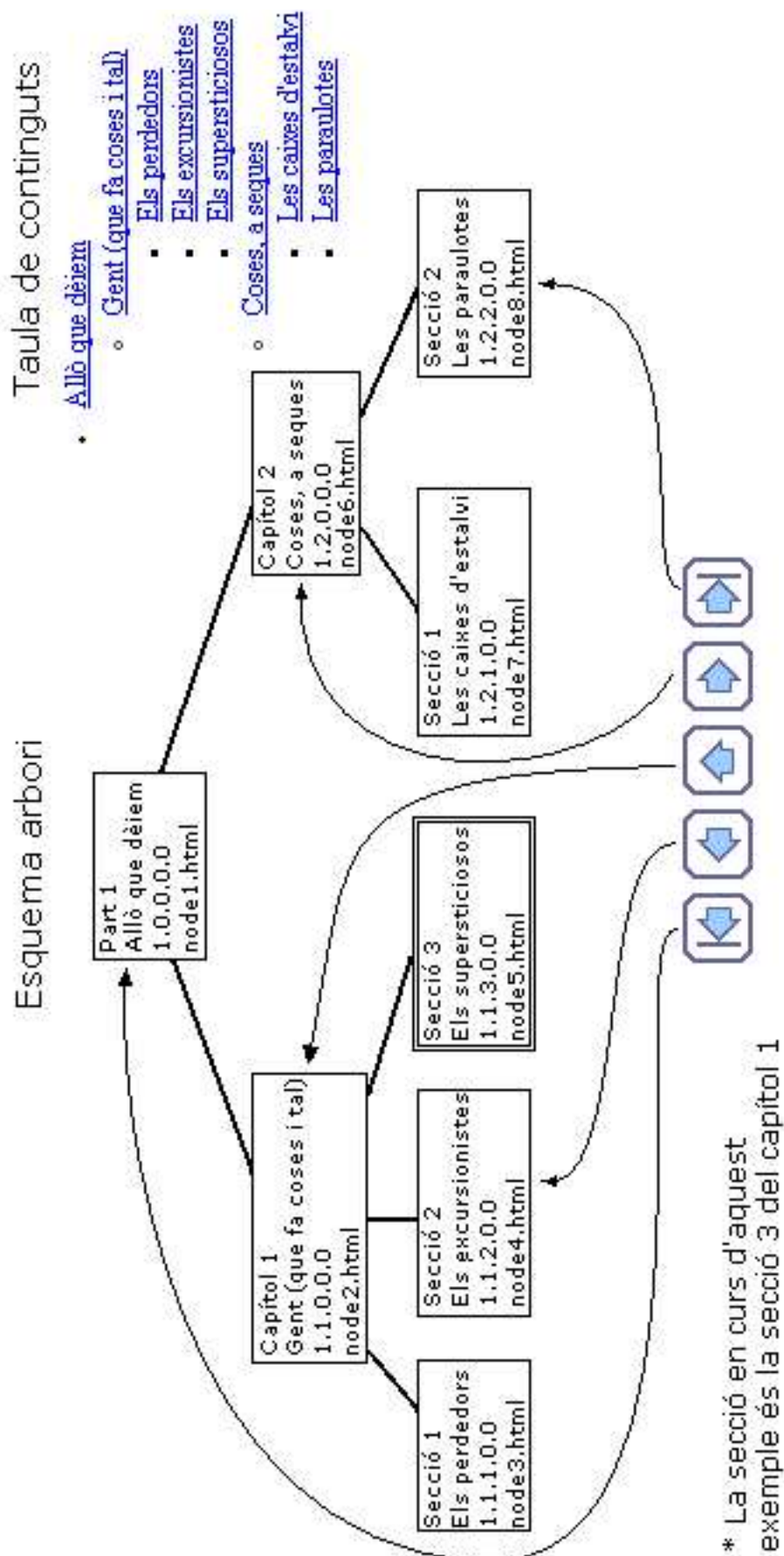


Figura 6.8: Exemple de navegació

Capítol 7

Proves fetes

En les pròximes pàgines es mostren diferents proves realitzades amb el traductor. S'han capturat diversos documents carregats en varis navegadors webs o browsers, cada apartat mostra diferents proves sobre un element concret dels que s'han explicat en temes anteriors.

1. Taula comparativa de traductors
2. Fragments matemàtics.
3. Llistes normals i llistes enumerades.
4. Taules de continguts, llistes de figures i llistes de taules.
5. Entorn *tabbing*
6. Figures
7. Títol creat mitjançant `\maketitle`
8. Fulla d'estil aplicada
9. Entorn amb marcs
10. Entorn *rawhtml*

★ Traductor proposat	LATEX2HTML	HeVeA	TtH	TeX4ht	GloboTeX *
Requereix compilador LaTeX	X			X	X
Requereix software extern com GS o Netpbm	X			X	X
Ràpida traducció		X	X		X
La traducció pot arribar a sér molt lenta	X			X	X
Altament configurable	X			X	X
Genera HTML de càrrega ràpida		X	X		
Fàcil d'instal·lar i usar			X		X
Traducció de matemàtiques a mode text		X	X		
Configurable utilitzant fitxer de configuració		X			X
Suport multilinguatge	X			X	
Ofereix bons resultats	X	X	X	X	X
Pot traduir qualsevol matemàtica	X			X	

Figura 7.1: Taula comparativa de traductors de L^AT_EX a HTML

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

Bibliografia

- [1] Valiente i Feruglio, Gabriel (1997). *Composició de textos científics con LaTeX* Edicions UPC, Barcelona, 1997.
- [2] M. E Lesk and E. Schmidt *Lex - A Lexical Analyzer Generator*
- [3] Paxson, Vern (1995). *Flex, version 2.5 A fast scanner generator*
- [4] C. Johnson, Stephen *Yacc: Yet Another Compiler-Compiler*
- [5] Donnelly, Charles & Richard Stallman (1995). *Bison The YACC-compatible Parser Generator Version 1.25*
- [6] W3C (1999). *HTML 4.01 Specification*
- [7] Drakos, Nikos (1999). *Text To Hypertext Conversion with LaTeX2HTML*
- [8] H. Hutchinson, Ian (2003) *TtH: The TEX to HTML translator*
- [9] Maltby, Gavin (1992). *An introduction to TeX and friends*
- [10] Gonzato, Guido (2002). *LaTeX for Word Processor Users*

Apèndix A

Manual d'aplicació per a l'usuari

En els pròxims apartats s'explica com instal·lar i utilitzar adequadament el traductor, així com els requeriments necessaris i la correcta utilització del fitxer de configuració.

A.1 Requeriments

Hardware El hardware mínim necessari per a poder utilitzar el traductor ha de ser aquell que permeti almenys compilar el programa sense cap mena de problema. En aquest sentit, qualsevol màquina amb unes mínimes prestacions és suficient.

Software Per tal que el traductor de L^AT_EX a HTML funcioni en la seva forma més bàsica és necessari disposar únicament d'un compilador de C++ per a poder compilar i executar el programa. Disposar del fitxer binari del traductor permetrà només traduir documents L^AT_EX que no disposin de figures ni de fórmules matemàtiques i símbols especials.

El funcionament complet del traductor implica tenir instal·lat el següent software addicional:

- Qualsevol distribució de L^AT_EX instal·lada al sistema que permeti la compilació de documents a format **PostScript**.
- **GhostScript** amb les seves corresponents eines de conversió gràfiques instal·lades al sistema.

A.2 Instal·lació

El programa es compila de la següent forma:

```
g++ globo.cc -o globo
```

...o d'una altra forma depenent del compilador de C++ que s'utilitzi. El fitxer executable generat pot ser copiat en qualsevol directori, i dins d'aquest, s'hi han de copiar els scripts `latex2gif` i `ps2gif`, a més a més del directori de les icones del programa.

A.3 Utilització

El programa s'utilitza per mitjà de paràmetres. La traducció d'un fitxer es realitza de la següent forma:

```
.\globo fitxer
```

Aquesta operació crea un directori amb el mateix nom del fitxer i dins d'aquest hi guarda la traducció en HTML juntament amb els fitxers gràfics generats. Si l'usuari ha decidit utilitzar un panell de navegació gràfic aleshores es copien també les icones del panell de navegació al directori.

La configuració del fitxer de configuració es realitza de la següent forma:

```
.\globo -c
```

Aquesta operació permet ajustar les propietats del fitxer de configuració per mitjà d'una línia de comandes.

Per saber quina versió del traductor estem fent servir podem utilitzar el programa de la següent forma:

```
.\globo -v
```

A.4 Utilització del fitxer de configuració

El fitxer de configuració del programa és el fitxer *globo.cfg*. Aquest fitxer el crea automàticament el traductor amb una configuració per defecte, si el fitxer ja existeix, aleshores el traductor carrega la configuració definida en el fitxer a la memòria. El traductor busca i carrega *globo.cfg* en el directori des de on s'executa, per tant, es pot disposar d'un fitxer de configuració diferent per cada document que pretenguem traduir, sempre i quan aquests es trobin en rutes diferents.

El fitxer de configuració es pot ajustar per mitjà d'un sistema de línia de comandes que inclou el traductor. Per utilitzar aquesta línia de comandes, s'ha

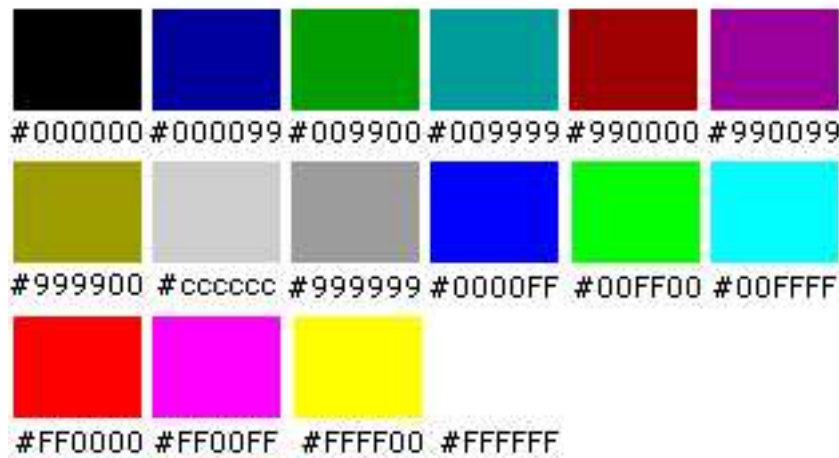


Figura A.1: Els 16 colors bàsics i la seva codificació en hexadecimal

d'arrancar el programa tal i com s'explica en l'apartat A.3. Tot seguit es detallen aquestes comandes.

titol Defineix un títol per a la finestra del navegador web o browser. S'utilitza de la següent forma:

```
titol títol
```

colorf Defineix el color de fons de tots els documents HTML que el traductor generarà. Els colors s'introdueixen en format hexadecimal, la figura A.1 mostra els 16 colors bàsics i la seva corresponent codificació. Aquesta comanda s'utilitza de la següent forma:

```
colorf #color
```

colorl Defineix un color per als enllaços. Els colors s'introdueixen en format hexadecimal, la figura A.1 mostra els 16 colors bàsics i la seva corresponent codificació. Aquesta comanda s'utilitza de la següent forma:

```
colorl #color
```

colorv Defineix un color per als enllaços ja visitats. Els colors s'introdueixen en format hexadecimal, la figura A.1 mostra els 16 colors bàsics i la seva corresponent codificació. Aquesta comanda s'utilitza de la següent forma:

```
colorv #color
```

colora Defineix un color per als enllaços actius. Els colors s'introdueixen en format hexadecimal, la figura A.1 mostra els 16 colors bàsics i la seva corresponent codificació. Aquesta comanda s'utilitza de la següent forma:

`colora #color`

css Defineix quin fitxer s'utilitzarà com a fulla d'estils per a la web. L'usuari haurà de copiar el fitxer `css` dins del directori del document on es troben els fitxers HTML generats. Si no s'especifica cap fitxer aleshores el document es presentarà en la seva forma original en HTML. Aquesta comanda s'utilitza de la següent forma:

`css fitxer`

refc Defineix el comportament dels enllaços de les referències creuades. Amb el paràmetre 1 les referències apunten a l'inici de l'apartat o entorn on s'ha inserit la macro *label*. Amb el paràmetre 0 les referències apunten just a la línia on s'ha inserit la macro *label*. Aquesta comanda s'utilitza de la següent forma:

`refc (1,0)`

nopart Defineix el format de l'estructuració general del document HTML. Amb el paràmetre 1 els apartats de tipus *part* no s'estructuren i per tant no apareixen a la taula de continguts, tot i que els continguts de les *parts* s'ajunen amb els continguts del nivell inferior. Amb el paràmetre 0 els apartats de tipus *part* si que s'estructuren. Aquesta comanda s'utilitza de la següent forma:

`nopart (1,0)`

nocapitol Defineix el format de l'estructuració general del document HTML. Amb el paràmetre 1 els apartats de tipus *capitol* no s'estructuren i per tant no apareixen a la taula de continguts, tot i que els continguts dels *capítols* s'ajunen amb els continguts del nivell inferior. Amb el paràmetre 0 els apartats de tipus *capitol* si que s'estructuren. Aquesta comanda s'utilitza de la següent forma:

`nocapitol (1,0)`

nosec Defineix el format de l'estructuració general del document HTML. Amb el paràmetre 1 els apartats de tipus *secció* no s'estructuren i per tant no apareixen a la taula de continguts, tot i que els continguts dels *secció* s'ajunen amb els continguts del nivell inferior. Amb el paràmetre 0 els apartats de tipus *secció* si que s'estructuren. Aquesta comanda s'utilitza de la següent forma:

`nosec (1,0)`

nosubsec Defineix el format de l'estructuració general del document HTML. Amb el paràmetre 1 els apartats de tipus *subsecció* no s'estructuren i per tant no apareixen a la taula de continguts, tot i que els continguts de les *subseccions* s'ajunten amb els continguts del nivell inferior. Amb el paràmetre 0 els apartats de tipus *subsecció* si que s'estructuren. Aquesta comanda s'utilitza de la següent forma:

nosubsec (1,0)

nosubsubsec Defineix el format de l'estructuració general del document HTML. Amb el paràmetre 1 els apartats de tipus *subsubsecció* no s'estructuren i per tant no apareixen a la taula de continguts, tot i que els continguts de les *subsubseccions* s'ajunten amb els continguts del nivell superior. Amb el paràmetre 0 els apartats de tipus *subsubsecció* si que s'estructuren. Aquesta comanda s'utilitza de la següent forma:

nosubsubsec (1,0)

panell Oculta o no el panell de navegació ens els documents HTML. Amb el paràmetre 1 el traductor no insereix el panell de navegació als fitxers HTML. Amb el paràmetre 0 el traductor insereix el panell de navegació en la part superior i inferior dels fitxers HTML. Aquesta comanda s'utilitza de la següent forma:

panell (1,0)

tpanell Permet escollir el tipus de panell de navegació. Amb el paràmetre 1 s'utilitza un panell de navegació en mode text. Amb el paràmetre 0 s'utilitza un panell de navegació gràfic basat en icones. Aquesta comanda s'utilitza de la següent forma:

tpanell (1,0)

menu Determina si s'utilitzarà o no el marc esquerra. Amb el paràmetre 1 el traductor no crea el marc esquerra i per tant el document HTML final no disposa de la taula de continguts a l'esquerra. Amb el paràmetre 0 si que es crea el marc esquerra. Aquesta comanda s'utilitza de la següent forma:

menu (1,0)

tmenu Especifica el tamany del marc esquerra en tant per cent. Aquesta comanda s'utilitza de la següent forma:

tmenu 0..100

barra Oculta o no la barra central de separació dels marcs. Amb el paràmetre 1 s'oculta la barra. Amb el paràmetre 0 no s'oculta la barra. Aquesta comanda s'utilitza de la següent forma:

`barra (1,0)`

path Especifica la ruta on es troben les icones del programa. Aquesta comanda s'utilitza de la següent forma:

`path ruta`

A.5 Detecció d'errors

Quan es detecta un error, normalment produït per l'incorrecte format del document L^AT_EX, es mostra la línia i el caràcter a partir del qual s'ha trobat l'error. El traductor contempla dos tipus d'errors: errors lleus que no alteren el transcurs de la traducció i errors greus que impliquen l'aturada de la traducció, donat que la traducció en curs és inconsistent. En el primer cas es mostra un avís indicant l'origen de l'error i es continua el procés, en el segon cas s'atura la traducció i es mostra també el motiu.

En els següents apartats s'explica els errors que es poden produir en temps de traducció i els missatges que mostra el traductor quan es produeixen.

A.5.1 Errors d'estructura

Els errors d'estructura es produeixen quan l'analitzador sintàctic no és capaç de reconèixer la sintaxi d'un tram determinat del document. Un exemple típic d'aquest tipus d'errors és quan l'usuari s'oblida de tancar un entorn que havia obert amb anterioritat. Moltes vegades aquests tipus d'errors són ocasionats per una mala interpretació del traductor davant de macros que desconeix, per tant, resoldre aquest tipus d'errors consisteix en corregir anteriors errors que s'han produït.

`Error-> ínia 124: Error de sintaxi. Revisa l'estructura`

La línia del document que causa un error d'aquest tipus no acostuma a ser mai la ínia que indica el traductor, ja que l'origen de l'error sempre es produeix en línies anteriors. Els errors d'estructura es consideren errors greus i per tant, impliquen l'aturada total del procés de traducció.

A.5.2 Macros no reconegudes

És el tipus d'error més habitual i sovint desencadena futurs errors de tipus estructural. El traductor contempla uns criteris predefinitos per afrontar aquest tipus d'errors amb la màxima fiabilitat possible. Els errors d'aquesta classe es consideren errors lleus i permeten continuar el procés de traducció amb normalitat, tot i que la traducció esdevé susceptible de detectar errors d'estructura.

Avís-> línia 5:78: Macro \newpage desconeguda pel traductor

A.5.3 Avisos

Hi ha situacions que no es consideren errors tot i ser fruit d'una incorrecta composició del document. Aquests casos es contemplen com avisos per a l'usuari, informant-lo del motiu. Els exemples més habituals són les referències a etiquetes inexistents dins del document.

S'ha trobat una referència a una etiqueta inexistent

Apèndix B

Conclusió i propostes per al futur

Per concloure aquest treball donaré el meu punt de vista sobre diferents aspectes relacionats amb tot el que he fet i com ho he fet.

La idea de desenvolupar un traductor de \LaTeX a HTML havia sorgit inicialment com una proposta de treball plantejada pel tutor, en aquest sentit, el fet que el projecte podia ser realment una cosa útil, és a dir, m'oferia la possibilitat de dissenyar i implementar una eina útil a la pràctica, em va atraure més que la idea de poder aprendre nous coneixements i conceptes que fins ara desconeixia.

Des del meu punt de vista personal, he valorat molt positivament els resultats del projecte, que en tot moment han estat conseqüència d'un línia de treball ordenada, rigorosa, i sobretot, meticulosa en el sentit de que la improvisació ha estat nul·la en tot el cicle del projecte, i per tant, s'ha complert estrictament l'organigrama de treball que m'havia plantejat desde el primer moment.

La primera fase del projecte, i possiblement la més pesada i feixuga, ha estat la fase de recerca i recopilació d'informació referent al sistema \LaTeX , tema amb el que no estava gaire familiaritzat. En aquesta primera fase era molt important concretar i deixar ben clar quines macros, entorns i aspectes del \LaTeX habien de poder ser traduïts pel traductor. La segona fase ha consistit en decidir com s'havia de realitzar la traducció per cada un dels elements concretats a la primera fase. La tercera fase del projecte consistia en especificar i implementar la gramàtica mitjançant **Flex++** i **Bison++**, i a la quarta, i última fase, ja només quedava implementar el conjunt.

L'objectiu inicial s'ha assolit, i crec que s'ha creat una eina útil, què pot ser utilitzada per estudiants, informàtics, i sobretot, per aquella gent que té especial interès en compondre i redactar documents ben fets. En aquest sentit, s'ha creat

una aplicació fàcil d'utilitzar, ràpida, eficient i sobretot que pot ser utilitzada per programadors que vulguin inserir còmodament un component traductor de \LaTeX a HTML als seus programes implementats en C++.

A continuació es descriuen un seguit de propostes per a possibles futures versions del programa.

- Permetre la traducció de formulacions i simbologies matemàtiques en mode text.
- Dotar el traductor d'un sistema de precompilació que faci possible la traducció de documents \LaTeX encapsulats per mitjà de la macro *input*
- Aplicar la metodologia **Babel** al traductor, o el que és el mateix, dotar el traductor d'un sistema multilinguatge.
- Ampliar el nombre de macros que és capaç de traduir.
- Oferir nous sistemes de navegació a través del document HTML.
- Fer possible que el traductor pugui realitzar les conversions gràfiques de forma totalment independent, sense dependre d'un compilador \LaTeX ni de cap intèrpret de llenguatge **PostScript** com el **GhostScript**.

Les propostes anteriorment mencionades, obren les portes a futures línies de treball, les quals no han estat desenvolupades en aquest projecte ja que des de el principi no formaven part dels plans inicials establerts en aquest.

Apèndix C

Preguntes més freqüents

Aquest traductor és software lliure. Això què significa? Significa que aquest traductor és lliure en el sentit de que amb total llibertat es pot executar, copiar, distribuïr, estudiar, modificar i millorar.

Si aquest traductor és lliure, aleshores per què utilitza el format gràfic GIF que no ho és? Aquesta no ha estat una decisió personal, sinó que més aviat ha estat una decisió forçada per la situació. De tots els formats disponibles, aquest és l'únic que d'alguna manera o altra està fortament vinculat amb els formats gràfics usats en l'Hypertext. Tot i que una adient modificació en l'script ps2gif i latex2gif del traductor permetrà a l'usuari utilitzar la resta de formats disponibles del GhostScript i evitar així el format GIF.

És possible inserir codi HTML en un document \LaTeX i que aquest, un cop traduït, pugui ser interpretat pel navegador? Sí, és possible. El traductor pot tractar fragments de codi HTML de dos maneres: com a text o com a codi HTML en el sentit estricte de la paraula. En el primer cas, el fragment es visualitza en el navegador com si es tractés d'un text qualsevol. El segon cas implica usar l'entorn *rawhtml*, amb el qual al traductor no li cal traduïr el fragment de codi, ja que ja es tracta de codi HTML suposadament vàlid.

No m'agrada com ha quedat el document HTML generat. Què puc fer? Traduir un document \LaTeX a HTML no vol dir fer-lo més bonic. El resultat d'un document traduït mai tindrà el mateix aspecte tipogràfic que l'original, tanmateix es poden associar fulles d'estil al document final per tal que aquest adopti un aspecte més elegant. Això és possible mitjançant una configuració adequada del fitxer de configuració.

M'agradaria que les fórmules matemàtiques i símbols especials es generin en format textual enlloc de generar-se en format gràfic. Això

és possible? Lamentablement això no és possible, almenys en aquest traductor proposat. La gran majoria de formulacions i simbologies matemàtiques que es poden fer en \LaTeX no tenen representació en HTML, degut a que HTML és un llenguatge pobre en aquest sentit. La solució més còmoda i fiable és la traducció directa a format gràfic mitjançant GhostScript.

Puc utilitzar aquest traductor en un sistema Windows? Sí, pots utilitzar aquest traductor en Windows, en Linux i en qualsevol altre sistema operatiu que permeti compilar i executar programes implementats en C++.

He provat de traduir un document que he fet en \LaTeX , però el traductor s'ha quedat penjat i no mostra senyals de vida. He fet alguna cosa malament? Davant aquesta situació tot apunta a que l'analitzador lèxic s'ha quedat penjat. L'analitzador lèxic no tolera l'anàlisi de lletres accentuades ni de caràcters especials, la qual cosa vol dir que el document \LaTeX que s'ha provat de traduir està mal fet ja que algunes lletres accentuades i/o caràcters especials no s'han escrit en el format adequat. Revisa el document i torna a arrancar el traductor.

No entenc per què surten tants missatges d'error i avisos quan tradueixo un document. Això té alguna explicació? Sí, en té. El traductor proposat reconeix un subconjunt bàsic del conjunt total de macros de \LaTeX , per tant, no és d'extranyar que algunes macros no siguin reconegudes i produeixin errors o avisos en temps de traducció. Tot i així, el traductor ha estat implementat amb l'objectiu d'ignorar totes aquelles macros que desconeix i realitzar les traduccions el màxim fiables possible.

Alguns símbols especials que el traductor se suposa que reconeix no es veuen al navegador web. Com ho puc solucionar? Alguns navegadors interpreten únicament la codificació de caràcters ISO 8859-1, també coneguda com ISO Latin-1, que contempla el conjunt de 256 caràcters que corresponen a la codificació ASCII. Alguns símbols són traduïts a la codificació Unicode que correspon a una extensió de ISO Latin-1, i per tant, codificacions de més de 255 caràcters, caràcters de més de 8 bits. Es recomana utilitzar les últimes versions dels navegadors web ja que tenen en compte aquesta codificació.

La conversió de determinades imatges i/o figures ha perdut qualitat. Es pot solucionar? El traductor contempla únicament la traducció d'imatges a format GIF, és a dir, els fitxers gràfics generats tenen una profunditat de color de 256 colors (píxels de 8bits). Millorar la qualitat dels fitxers generats implica realitzar la conversió a un format gràfic de més de 256 colors, per a fer això s'ha de modificar adequadament els scripts *ps2gif* i *latex2gif*.

Puc utilitzar el traductor des de qualsevol directori del sistema si la seva ruta està carregada al PATH? Si, es pot, però aleshores s'ha de configurar el fitxer de configuració i indicar on es troba la carpeta de les icones i també s'ha de carregar al PATH la ruta dels scripts latex2gif i ps2gif. Tot i això, s'ha d'ajustar la configuració dels scripts esmentats per tal que funcionin correctament.

On puc trobar actualitzacions i millores del programa? Les actualitzacions, millores i altres temes relacionats amb el traductor ho trobaràs a la web del seu autor a l'apartat **GloboTeX - LaTeX to HTML translator** a <http://arcana.ods.org>