

Mise au point d'une chaîne de conversion de
documents L^AT_EX en vue de leur diffusion sur
l'Internet

Youssef Wabi

septembre 2003

Rapport de stage (juin à juillet 2003)

A l'intention de :

M^{me} Sophie Tison et de M^{me} Marie-France Claerebout

Table des matières

1	Résumé	4
1.1	Mots-clés :	4
2	Introduction	5
3	Le projet Grisemine	6
4	La mise en forme des documents, les contraintes de diffusion et les conclusions de départ	6
5	L'environnement informatique et logiciel	8
5.1	dvipdfm	8
5.2	Acrobat Reader	8
6	Recherche Documentaire	9
6.1	Solutions existantes pour la production de pdf	9
6.1.1	ps2pdf	10
6.1.2	pdftex	10
6.2	Classes L ^A T _E X utilisées	11
6.2.1	la classe L ^A T _E X de l'université de Laval	11
7	Mise au point de notre classe	11
7.1	Modèle de la thèse française	12
7.2	La page de titre	12
7.3	L'inclusion des images et des schémas	14
7.4	Les liens hypertextes	15
8	Bilan et perspectives	16
8.1	Caractéristiques de la classe	16
8.2	Utilisation de la classe classeThese	16
8.3	Les points restés en suspens	17
8.3.1	encore les images	17
8.3.2	la bibliographie	17
8.3.3	l'index	17
8.4	Extensions	18
8.4.1	Mémoires de DEA	18
8.4.2	L ^A T _E X vers HTML	18
9	Conclusion	19

Table des figures

1	page de titre générique	13
---	-----------------------------------	----

1 Résumé

J'ai effectué après l'obtention de ma maîtrise d'informatique à l'université de Lille1 un stage au sein du service des thèses et de la littérature grise de la bibliothèque centrale de cette même université. Ce stage s'inscrivait dans le cadre du projet Grisemine, géré par cette section. Ce projet est relatif à la mise en ligne et à la consultation sur Internet des travaux de 3ème cycle (mémoires de thèses et de DEA). Le sujet de mon stage était : Mise au point d'une chaîne de conversion de documents \LaTeX en vue de leur diffusion sur l'Internet.

Je me suis dans un premier temps référé au travail d'un autre stagiaire dont je prenait la suite dans ce projet et ai donc logiquement inscrit mon travail dans la suite directe du sien en repartant des conclusions auxquelles il était parvenu. Après une phase de recherche documentaire qui m'a permis de comparer les diverses méthodes et outils disponibles pour la production de fichiers pdf à partir de fichiers sources \LaTeX , j'ai élaboré une classe \LaTeX qui permettait de structurer le document source afin d'obtenir le résultat escompté. La mise au point de cette classe a nécessité une attention particulière en ce qui concerne la page de titre, les images et aussi les liens hypertextes.

1.1 Mots-clés :

- Grisemine,
- thèses françaises,
- \LaTeX ,
- dvipdfm.

2 Introduction

Le stage que j'ai effectué de juin à juillet 2003 au sein du service des thèses et de la littérature grise de la bibliothèque centrale de l'université de Lille1 s'est déroulé sous la co-responsabilité de Mme Marie-France Claerebout, conservatrice du service et responsable du projet Grisemine (<http://bibliotheques.univ-lille1.fr/grisemine>) et de Mme Sophie Tison, professeur d'informatique directrice adjointe du LIFL (<http://www.lifl.fr>) et directrice de l'Ecole Doctorale Sciences Pour l'Ingénieur (SPI) de Lille1.

Le sujet de ce stage était : *Mise au point d'une chaîne de conversion de documents \LaTeX en vue de leur diffusion sur l'Internet*. L'objectif était de mettre au point un modèle de thèse \LaTeX dont la compilation et la transformation au format pdf soient conformes aux contraintes de diffusion du projet Grisemine. Les conclusions de ce stage doivent servir de base de proposition de mise en forme des thèses \LaTeX provenant de l'école doctorale SPI. C'est pourquoi une attention particulière a été portée au respect au mieux des usages en vigueur au sein des diverses composantes de l'école doctorale et aussi à la conformité du document final avec les directives ministérielles de présentation des thèses françaises.

Mon travail a commencé par une prise de connaissance du travail réalisé par Laurent Candillier qui est le stagiaire de l'équipe Grappa du LIFL dont je prenais la suite. J'ai par la suite effectué une recherche documentaire pour identifier les méthodes et outils adéquats pour réaliser le passage du \LaTeX au pdf en respectant les contraintes posées. Il est à signaler qu'une modification de ces contraintes aurait conduit à un résultat différent. Je suis ensuite passé à la phase de mise au point de la classe \LaTeX dont l'utilisation doit permettre la production de fichiers conformes au résultat escompté par les bibliothécaires.

Dans une première partie, je présente le projet Grisemine et son environnement, les conclusions et les contraintes à partir desquelles j'ai commencé mon propre travail, l'installation informatique utilisée et les installations logicielles nécessaires. Le deuxième chapitre décrit les résultats de la recherche documentaire préalable au travail proprement dit, puis le travail que j'ai réalisé et ses résultats. La troisième partie explicite les différentes extensions possibles et souhaitables du travail fait et resitue celui-ci dans le cadre du projet Grisemine.

3 Le projet Grisemine

Grisemine est un projet de bibliothèque numérique de littérature grise francophone à diffusion internationale.

Grisemine met à disposition sur un serveur de l'université (<http://bibliotheques.univ-lille1.fr/grisemine>) des mémoires de thèses ou des extraits, des mémoires de DEA, des comptes-rendus de séminaires et colloques scientifiques, des cours, ... Ces documents sont consultables gratuitement et par tous. Ce projet en est à une première version qui est amenée à évoluer vers une collaboration avec d'autres acteurs francophones impliqués dans la recherche et/ou l'enseignement.

A l'heure actuelle, le serveur regroupe plus de 600 documents consultables. Il s'enrichit mensuellement de nouveaux textes. Le format pdf est celui retenu pour permettre la lecture en ligne des documents.

Grisemine se veut cohérent avec le projet national de diffusion électronique des thèses qui se met en place à l'initiative du ministère de l'Education Nationale et sous la supervision de la Sous-Direction des Bibliothèques et de la Documentation (SDBD) (<http://www.sup.adc.education.fr/bib/>). Ce service a mis au point un modèle de thèses, afin de faciliter la rédaction des thèses et leur diffusion.

Dans le cadre de ce stage, la structuration proposée pour les thèses doit donc être cohérente avec les recommandations ministérielles. La SDBD développe actuellement un dispositif de diffusion de thèses par voie électronique. Celui-ci prévoit une conversion du format natif de rédaction vers un format balisé d'archivage (SGML/XML). Cette partie du projet de diffusion sera coordonnée au plan national. La 2^{de} étape qui consiste en la diffusion proprement dite (aux formats PDF,HTML,SGML) est laissée à la charge des bibliothèques.

4 La mise en forme des documents, les contraintes de diffusion et les conclusions de départ

Les matériaux présentés sur Grisemine parviennent indifféremment au service des thèses sous le format Word et \LaTeX . Ils sont diffusés en pdf avec une évolutivité prévue vers l'HTML. Pour les documents au format \LaTeX , essentiellement des thèses et des mémoires, le fichier pdf obtenu est souvent de piètre qualité. Cette qualité souvent insatisfaisante pour une mise en ligne sur Internet a montré le besoin de mettre au point des outils de base en \LaTeX dont l'utilisation permettrait d'obtenir in fine des fichiers pdf de bonne qualité.

Les contraintes de diffusion sur ce projet sont peu nombreuses. Il fallait que les mémoires déposés sur le serveur disposent d'un menu en signets complètement renseigné (chapitres, sous-chapitres, annexes) et que celui-ci s'affiche automatiquement lors de l'ouverture du fichier avec Acrobat Reader. Le texte du mémoire devait quant à lui s'afficher en "Pleine Largeur" et à la taille 100% et être parfaitement lisible.

Le premier écueil rencontré par plusieurs thèses est celui de la lisibilité. Celle-ci est jugée par rapport à l'affichage dans Acrobat Reader, qui est le visualisateur pdf d'Adobe, très répandu et aussi très facile d'installation. Le format pdf possède quant à lui d'indéniables atouts par rapport aux formats dvi et postscript. Le format dvi utilise des polices de caractères standard intégré dans L^AT_EX. Il n'est pas adapté à une diffusion grand public, vu qu'il est improbable que les utilisateurs de GriseMine prendraient la peine d'installer une distribution de L^AT_EX juste pour pouvoir lire des fichiers dvi. Quant au format ps, il ne possède pas les fonctionnalités avancées du pdf : lien hypertextes, tables des matières en signets, renvois intra-documents. Un autre argument en faveur du pdf est qu'il est assez simple de produire du pdf à partir des fichiers Word.

L'année dernière, Laurent Candillier, doctorant au sein de l'équipe Grappa du Lifl avait fait un stage destiné à détecter les sources éventuelles dans les fichiers L^AT_EX d'origine de cette pauvre qualité de fichiers pdf finaux. L'objectif était de pouvoir suggérer aux auteurs dont les documents poseraient éventuellement problème d'adopter quelques unes de ces recommandations afin que le document pdf obtenu correspondent aux attentes des responsables du projet.

Les conclusions auxquelles il est parvenu sont les suivantes. Les instructions de compilation ci-dessous en début de fichier constituent une spécification minimale.

```
\documentclass[dvipdfm]{report}

\usepackage[frenchb]{babel}
\usepackage[dvipdfm]{hyperref}
\usepackage[dvips]{graphicx}
\usepackage[usenames,dvipsnames]{color}
\usepackage[latin1]{inputenc}
\usepackage{graphicx}
\usepackage{isolatin1}
```

Il recommande l'utilisation de *dvipdfm* pour obtenir le fichier pdf à partir du fichier dvi créé à la compilation du source L^AT_EX. *dvipdfm* doit donc en cela remplacer les 2 commandes *dvips* et *ps2pdf*. Dans l'optique de la production

de fichiers pdf en bout de course, il suggère aussi de préférer les paquetages adaptés au pdf plutôt que ceux adaptés au format postscript au début du fichier \LaTeX . C'est le cas par exemple du paquetage fontenc avec l'option T1. Il ne faut pas inclure l'instruction : `\usepackage[T1]{fontenc}` ou alors changer l'option T1 en OT1 ou OT4. Afin de pouvoir visualiser un fichier pdf qui s'affiche avec des signets reprenant la table des matières il faut inclure le module `hyperref`. A noter aussi que l'utilisation du type `article` en lieu et place de `report` est aussi valable.

5 L'environnement informatique et logiciel

A partir de ces conclusions et des contraintes de diffusion, je devais commencer mon travail. J'avais à cet effet à ma disposition un PC dans un bureau au sein du service des thèses de la bibliothèque et je pouvais à partir de là accéder aux diverses ressources documentaires d'un serveur appelé pcbu-grisemine et qui regroupe les documents en cours de traitements avant leur diffusion.

Après avoir installé Linux (Mandrake 8.1 pour le n° de version) à côté du système Windows qui tournait déjà sur le PC, j'ai procédé à l'installation des logiciels nécessaires pour mon travail. J'ai donc installé `dvipdfm` et Acrobat Reader.

5.1 `dvipdfm`

Il s'agit d'un logiciel qui permet de produire du pdf à partir d'un fichier `dvi`, celui-ci étant le résultat de la compilation du fichier source \LaTeX . `Dvipdfm` a été mis au point et est maintenu par Mark A. Wicks et est disponible en format rpm pour Linux sur plusieurs sites web. Je l'ai pour ma part téléchargé du serveur ftp de l'université Paris 8 : (<ftp://ftp.cs.univ-paris8.fr/pub/linux/distributions/mandrake/>). `Dvipdfm` s'utilise comme suit : `dvipdfm fichier.dvi` . Cette commande crée `fichier.pdf` qui est le pdf voulu.

5.2 Acrobat Reader

C'est bien entendu le logiciel de la société Adode (<http://www.adobe.com>) qui permet de visualiser les fichiers pdf. Il est gratuit, d'installation facile, supporté par plusieurs plateformes et très répandu. C'est cette facilité d'emploi jointe à sa large utilisation par les utilisateurs finaux (souvent sous Windows avec Internet Explorer) qui a fait qu'il ait été retenu dans ce projet

comme visualisateur pdf, aux dépens d' autres logiciels parfois déjà inclus dans la version de Mandrake installée.

La version d'Acrobat reader pour Linux utilisée dans ce stage est celle disponible sur le site gurulabs.com (<http://www.gurulabs.com/download.html>). Elle se présente sous la forme de 2 paquetages rpm :

- `acroread-<version-number>.i386.rpm`
- `acroread-plugin-<version-number>.i386.rpm`

Après l'installation du logiciel il faut le configurer afin de pouvoir bénéficier des fonctionnalités hypertextes du pdf. C'est à dire qu'il faut indiquer à Acroread quel navigateur web utiliser pour ouvrir une page web sur l'adresse de laquelle on vient de cliquer dans le document pdf. Dans le cas d'un clic sur une adresse email et de l'utilisation par l'utilisateur de Netscape comme navigateur et serveur de mail, c'est l'application Netscape Composer qui doit se lancer. Cette configuration du navigateur est accessible à partir de l'onglet Edit -> Preferences -> Weblink.

Pour résumer, la production des fichiers pdf à partir des sources L^AT_EX pendant ce stage se faisait de la manière suivante :

```
latex fichier.tex; latex fichier.tex; dvi2pdf fichier.dvi
```

Les 2 compilations sont nécessaires pour la génération correcte de la table des matières. Le fichier dvi reste bien sûr disponible (dvi2pdf ne l'efface pas et ne le transforme pas non plus) et les personnes qui restent attachées à la mise en ligne sur leurs pages web personnelles ou à la diffusion sur certains circuits de leurs documents en postscript pourront toujours créer les documents adhoc.

6 Recherche Documentaire

J'ai au cours de cette première étape de mon travail examiné plusieurs moyens de produire du pdf en partant de L^AT_EX afin de mettre à jour les conclusions de Laurent Candillier. J'ai aussi recherché un modèle de classe L^AT_EX pour les thèses sur lequel j'aurais pu m'appuyer pour commencer mes développements.

6.1 Solutions existantes pour la production de pdf

Il existe plusieurs de ces solutions. J'ai examiné (de plus ou moins près) dvi2pdf (déjà vu), ps2pdf et pdftex.

6.1.1 ps2pdf

ps2pdf est une application incluse dans la distribution Ghostscript et dont le principe est de produire du pdf en utilisant les options adéquates. Cependant, la qualité du résultat obtenue laisse parfois (voire souvent) à désirer. David Monniaux, chercheur au CNRS et auteur du très instructif "Comment faire du pdf avec L^AT_EX" (<http://www.di.ens.fr/~monniaux/download/pdf-en-latex.pdf>) en donne l'explication et la solution que je rapporte ici. Selon lui, il ne faut utiliser ps2pdf qu'avec une version de Ghostscript supérieure ou égale à la 6.0 . Les versions antérieures ne savent pas inclure dans le fichier pdf certains types de police et insèrent à la place des caractères écrits avec ces polices des images bitmap, qu'Acrobat Reader affiche très mal. Dans le cadre de notre projet, c'est essentiellement le confort de lecture des utilisateurs du site grisemine qui nous intéresse et il aurait fallu être sûr de pouvoir éventuellement demander aux rédacteurs d'utiliser une version "convenable" de Ghostscript pour avoir un résultat satisfaisant. Je considère cette solution comme "impraticable" en laboratoire de recherche où les chercheurs ont déjà leurs machines d'installées avec leurs logiciels dont certains sont disponibles via le réseau. Je crois qu'il n'est pas raisonnable de les amener à changer de version de logiciel juste pour satisfaire nos contraintes de diffusion de leurs travaux, qu'ils sont d'ailleurs libres de nous laisser diffuser ou non. Je ne me suis donc pas appuyé sur la solution ps2pdf de Ghostscript pour mon travail.

6.1.2 pdftex

pdftex est une application qui permet de créer un fichier pdf directement à partir d'un fichier L^AT_EX. Il présente l'inconvénient de ne pas passer par le format dvi. Pour obtenir un fichier postscript à partir du résultat de pdftex, il faudrait en fait utiliser l'application inverse de ps2pdf à savoir pdf2ps. Il est probable que le fichier pdf obtenu avec les instructions adéquates pour le pdf et pdftex ne conviennent plus pour le format ps et on retombe sur les mêmes écueils que pour ps2pdf. D'autre part pdflatex nécessite l'inclusion dans le fichier source L^AT_EX de commandes qui lui sont spécifiques et la compilation se fait par l'appel de la commande : `pdftex fichier.tex` et non pas de la commande `latex fichier.tex`. Je n'ai pas retenu cet outil car son utilisation introduit une nouveauté par rapport aux usages des rédacteurs. En conclusion de cette recherche logicielle, j'ai décidé de conserver l'outil dvipdfm déjà utilisé par Laurent Candillier et aussi de reprendre toutes ses autres conclusions comme point de départ de mon travail.

6.2 Classes L^AT_EX utilisées

J'ai recherché des exemples de classe de thèse et ai trouvé 2 documents qui m'ont été très utiles dans mon travail. Ce sont d'une part un exemple de classe L^AT_EX pour thèse francophone de l'université de Laval au Québec et d'autre part un exemple de classe L^AT_EX de mise en forme de rapport de stage.

6.2.1 la classe L^AT_EX de l'université de Laval

La Faculté des études supérieures de l'université de Laval au Québec a développé des outils pour les traitements de texte Word et L^AT_EX dans l'optique du dépôt et de la diffusion électronique des mémoires et des thèses. Ces 2 outils sont une feuille de style pour le format Word et une classe pour le format L^AT_EX et ils implémentent tous les 2 les règles et conventions typographiques de la Faculté. La classe L^AT_EX en question est disponible à l'url suivante : <http://theses.ulaval.ca/formation.html> Compte tenu de la similitude entre le projet Grisemine et le projet de l'université de Laval, je me suis beaucoup appuyé sur le travail de ces derniers. La perspective d'utiliser au final un outil assez ressemblant (aux conventions d'écriture et de présentation près) est très intéressante et est susceptible de pouvoir créer des échanges, ce qui est aussi un objectif des promoteurs de Grisemine. La classe L^AT_EX de Laval m'a donc servi de point de départ en plus du travail déjà accompli par Laurent Candillier. J'ai aussi utilisé la classe L^AT_EX pour l'écriture de rapport de stage mise au point par Virginie Quesnay (<http://virginie.quesnay.free.fr/>), une étudiante en IUP Génie des Systèmes d'Information.

7 Mise au point de notre classe

La première chose dont j'ai eu besoin était de connaître la norme de mise en page des thèses françaises. Une fois que ce fut fait, j'ai travaillé à la mise en forme de certains points particuliers comme la page de titre, les images et les liens hypertextes.

Pour illustrer les différents aspects de ce travail, j'ai repris comme fichier d'exemple un de mes rapports de stage et lui ait donné la structure d'une thèse. Le stage en question a été effectué l'année dernière au Lifi sous la responsabilité entre autre de Mme Tison.

7.1 Modèle de la thèse française

Le modèle de présentation de la thèse française que m'a communiqué Mme Claerebout est structuré comme suit avec dans l'ordre :

- la page de titre
- la dédicace
- les remerciements
- le résumé en français avec comme sous-paragraphe les mots-clés
- le résumé en anglais avec les mots-clés en anglais
- la table des matières
- la liste des tableaux
- la liste des figures
- l'introduction
- les différents chapitres
- la conclusion
- la bibliographie
- les annexes

7.2 La page de titre

Celle que nous proposons commence avec tout en haut au centre le logo de l'université, juste en dessous son nom complet et en dessous le nom de la composante (UFR ou Département) où s'est déroulée la thèse. La figure suivante donne l'allure de la page complète que l'on obtient quand on ne remplace pas dans le fichier d'exemple les données de la page de titre par de véritables données. Sa mauvaise qualité est due au fait que c'est une capture d'écran. Pour renseigner les différents champs par leurs bonnes valeurs, le rédacteur utilisant la classe n'aura qu'à utiliser le fichier template où il déclare dans son préambule faire appel à cette classe et écrire en lieu et place des valeurs par défaut les valeurs adéquates. Pour ce qui est de la page de titre les données par défaut se présentent comme suit :

```
\documentclass[dvipdfm]{classeThese}

\begin{document}

\Composante{ [nom de la composante] }
\SousComposante{ [nom de la sous-composante] }
\NumeroOrdre{ [numero] }
\Discipline{ [nom de la discipline]}
\PreNomNomMaj{[Pr\'ENOM NOM] }
\PreNomNomMin{[Pr\'enom Nom] }
```



Université des Sciences et Technologies de Lille

[nom de la composante]

[nom de la sous-composante]

N° d'ordre: [Numero]

THÈSE

Pour l'obtention du grade de
Docteur de l'Université des Sciences et Technologies de Lille

Discipline : [nom de la discipline]

présentée et soutenue publiquement par :
[PRÉNOM NOM]

[date de la soutenance]

[titre de la these]

Directeur de thèse :
[prenom nom directeur de these]

Jury :		
prenm nom président(e) de jury	titre universitaire	Président
prenm nom membre du jury	titre universitaire	rapporteur ou examinateur
prenm nom membre du jury	titre universitaire	rapporteur ou examinateur
prenm nom membre du jury	titre universitaire	rapporteur ou examinateur
prenm nom autre membre	titre universitaire	rapporteur ou examinateur

FIG. 1 – page de titre générique

```

\DateSoutenance{ [date de la soutenance] }
\titre{ [titre de la these] }
\PreNomNomDirecteurThese{ [prenom nom directeur de these] }

\Jury{
\begin{tabular}[b]{l}
prenom nom pr\’esident(e) de jury & titre universitaire & Pr\’esident & \\\
prenom nom membre du jury & titre universitaire & rapporteur ou examinateur & \\\
prenom nom membre du jury & titre universitaire & rapporteur ou examinateur & \\\
prenom nom membre du jury & titre universitaire & rapporteur ou examinateur & \\\
prenom nom autre membre & titre universitaire & rapporteur ou examinateur & \\\
\end{tabular}
}
\maketitle

```

7.3 L’inclusion des images et des schémas

Certaines images incluses dans les fichiers pdf obtenues à partir de source \LaTeX ne sont pas toujours bien nettes (pour ne pas dire carrément floues). Il fallait régler ce problème et pouvoir faire des propositions permettant de tenir compte de la grande variété des formats d’images disponibles.

Le paquetage de gestion des images qu’on utilise est `graphicx` avec le pilote `dvipdfm`. Ceci nous donne dans le préambule la ligne :

```
\RequirePackage[dvipdfm]{color,graphicx}
```

Pour que l’environnement `\figure` puisse gérer les différents formats d’images que l’auteur serait amené à inclure, il faut déclarer toutes les extensions à considérer comme suit :

```
\DeclareGraphicsExtensions{.png,.jpg,.mps,.pdf,.gif,.tif}
```

Il est aussi recommandé au rédacteur de ne pas mettre d’extension au nom de ses images quand il utilise la commande `\includegraphics`. Dans ce cas le système se charge lui-même de retrouver l’image correspondante dans le chemin d’accès aux images dans l’ordre de déclaration des extensions.

Je dois cependant préciser que pour l’inclusion des images, j’ai été confronté à un écueil particulier : celui de la bounding box. Pour toutes les images non ps ou non eps, \LaTeX me réclamait la bounding box. La bounding box est la plus petite boîte dans laquelle s’insère une figure. Les dimensions de la bounding box attachée à une figure doivent être indiquées en paramètre de la commande `\includegraphics` de la manière suivante :

```
\includegraphics[bb= LLx Lly URx URy]{./images/laFigure}.
```

Ces données correspondent à :

- LLx : coordonnée horizontale du coin inférieur gauche (Left Low x pour

- s'en souvenir)
- LLy : coordonnée verticale du coin inférieur gauche
- URx : coordonnée horizontale du coin supérieur droit
- URy : coordonnée verticale du coin supérieur droit

Ces dimensions sont calculées en points postscript à partir du coin inférieur gauche de la page et on a l'équivalence 72 points = 1 inch = 25.4 mm. Je n'ai pas trouvé le moyen de me passer de la spécification de ces paramètres pour l'inclusion des images. La seule solution acceptable que j'ai pu identifier est celle qui consiste à convertir une image jpg, png, ... en ps grâce à Gimp, par exemple, puis d'ouvrir l'image ps obtenue (souvent de mauvaise qualité) dans Ghostview. En déplaçant la souris sur les coins inférieur gauche et supérieur droit, on peut lire les coordonnées x-y de la bounding box. Il suffit alors de reprendre ces chiffres lors de l'inclusion de l'image originale non ps dans le source \LaTeX . Cette solution me paraît acceptable, dans la mesure où on n'inclut que les images originales non ps et que souvent celles-ci ont une taille (en octets) inférieure aux images ps équivalentes. On obtient ainsi une baisse de la taille du mémoire final, ce qui aux yeux des bibliothécaires est un paramètre important.

7.4 Les liens hypertextes

Ils sont activés par l'utilisation du paquetage hyperref, qu'il faut donc aussi charger dans le préambule mais en dernier. Le paquetage hyperref permet l'utilisation des différentes références du document \LaTeX : table des matières, index, notes de bas de page ... Il est paramétrable avec plusieurs options de deux manières : soit dès d'appel avec les options entre crochets, soit avec la commande `\hypersetup`. J'ai préféré la 1ère solution, vu que j'utilise la 2ème pour les méta-données.

Une attention particulière était portée dans ce travail à l'activation des liens externes, c'est çà dire vers des pages web ou alors vers des emails. On peut utiliser la commande : `\href`.

`\href` s'emploie comme suit :

`\href{url où doit renvoyer le clic}{texte sur lequel on clique}`.

L'url doit être complète, c'est à dire avec son protocole en entête. Une commande du genre : `\href{www.lifl.fr}{lifl}` ne marche pas ! Il faut obligatoirement mettre `http://` devant. De même pour les adresses emails il faut indiquer le protocole `mailto`, comme ceci :

`email : \href{mailto:wabi@fil.univ-lille1.fr}{wabi@fil.univ-lille1.fr}`.

Ce qui donne : wabi@fil.univ-lille1.fr.

En fait href se comporte comme n'importe quelle commande qui serait définie de la manière suivante :

`\newcommand{\<nom de la commande>}[2]{\htmladdnormallink{#2}{#1}}`.
Il suffirait alors de faire : `\<nom de la commande>{url}{texte à cliquer}`
pour avoir le même résultat.

8 Bilan et perspectives

Les caractéristiques et le mode d'emploi de la classe finale sont rappelés dans le paragraphe suivant. Cette classe est susceptible d'être améliorée en raison des problèmes que je n'ai pas su régler ou que je n'ai pas abordé. Cependant cette classe peut servir de base pour de futurs développements dans le cadre du projet Grisemine.

8.1 Caractéristiques de la classe

La classe `classeThese` permet donc de mettre en forme une thèse de l'université de Lille1. `classeThese` commence par une page de titre reprenant toutes les informations requises par la nomenclature des thèses françaises. Elle définit un environnement pour la dédicace et les remerciements. Les mots-clés en français et en anglais figurant sur les pages des résumés ne se retrouvent pas dans les méta-données du document pdf final. Les signets à gauche affichent la table des matières à partir du résumé en français et quant à la table de matières elle-même, elle comporte des liens hypertextes non colorés. Il n'y a que le changement de forme de la souris pour vous indiquer la présence du lien. Les pages sont numérotées à partir de la page de titre jusqu'à l'introduction (toutes les 2 exclues) en chiffres romains et à partir de là en chiffres arabes.

8.2 Utilisation de la classe `classeThese`

Pour utiliser la classe il faut récupérer le fichier `classeThese.cls` et `templateThese.tex`. Il est prévu qu'il soit téléchargeable depuis une page web sur l'intranet de la bibliothèque. Il faut ensuite créer un fichier principal dans le répertoire où l'on a téléchargé le fichier et un répertoire nommé `images` qui contiendra toutes les images utilisées dans les fichiers \LaTeX . C'est surtout une commodité car on n'a pas défini de commande spéciale pour l'insertion de manière uniforme des images. Il faut avant que le mode opératoire suggéré soit définitivement retenu.

Le rédacteur peut indifféremment écrire son texte dans le fichier principal `templateThese.tex` à la place voulue soit l'inclure en utilisant la commande

`\input` comme suit : `\input{chapitre1.tex}`. Idéalement la première solution devrait être utilisée pour les pages de remerciements et dédicace et aussi pour les transitions entre chapitres s’il y a lieu.

8.3 Les points restés en suspens

8.3.1 encore les images

Le premier point à régler est à mon avis l’inclusion des images. La nécessité de devoir préciser les coordonnées de la bounding box n’est pas du tout satisfaisante. Je pense qu’il doit y avoir une solution mais il me reste encore (à moi ou à un autre) à trouver laquelle. Pour se dispenser de la “corvée” que représente le fait de devoir préciser les coordonnées de la bounding box pour chaque image non ps, on aurait pu utiliser pdftex au lieu de dvi_{ps} et alors convertir les éventuelles images ps en un autre format, mais réflexion faite, je pense que les 2 solutions se valent ...

8.3.2 la bibliographie

La bibliographie est gérée par l’environnement `\thebibliography` avec les commandes `\bibitem{marque}` ouvrage cité et `\cite{marque}` pour y faire un renvoi. L’utilisation possible de l’outil BiBTeX n’a pas été étudiée.

8.3.3 l’index

La classe ne gère pas les index. Dans le cadre d’une thèse sa présence ne se justifie pas forcément et au vu des thèses reçues au service des thèses, il apparait que cette pratique est très rare. Il n’en reste pas moins possible de mettre un index, ce qui peut arriver si l’on utilise cette classe pour rédiger un article scientifique. Dans ce cas il faut insérer les déclarations ad hoc dans le fichier template :

- `\usepackage{makeidx}` et `\makeindex` dans le préambule
- `\printindex` juste avant le `\end{document}`

Il suffit alors de faire suivre chaque mot devant apparaitre dans l’index de la commande `index{mot-clef de l’index}`. La génération de l’index nécessite l’emploi de la commande : `makeindex <nom du fichier (sans extension)>` entre les 2 appels à la commande `latex`.

8.4 Extensions

8.4.1 Mémoires de DEA

La classe mise au point à partir de la nomenclature des thèses françaises peut s'étendre aux mémoires de DEA . Un exemple a été mis au point à partir du modèle proposé par Mr Beckermann, responsable du DEA de Mathématiques Appliquées pour sa formation. Ce modèle a été récupéré à l'adresse suivante : <http://ano.univ-lille1.fr/bb/deainfo/>.

Les fichiers classeThese et templateThese ont été modifiés pour permettre l'inclusion des logos et autres éléments distinctifs de la composante à partir du template et non plus de la classe.

Cette solution est particulièrement adaptée aux composantes qui veulent signaler leur identité (logo) sur les documents produits par leurs membres. Les modèles classeThese et templateThese restent à la disposition de ceux qui n'ont pas encore développé de page de titre spécifique ou qui ne souhaitent tout simplement pas le faire. Le plus important est que les fichiers obtenus à partir de ces 2 démarches respectent encore les contraintes de diffusion du projet Grisemine.

8.4.2 \LaTeX vers HTML

La transformation du source \LaTeX en fichier HTML est aussi à envisager. Il existe pour cela des outils comme LaTeX2HTML ou HEVEA.

Le traducteur LATEX2HTML a été écrit par Nikos Drakos de l'University of Leeds à Leeds (Grande-Bretagne) puis étendu et revu par Ross Moore de la Macquarie University à Sydney (Australie). Il est distribué gratuitement mais sans garantie. On peut cependant remarquer qu'il est déjà très utilisé sur Internet, pour la mise en ligne de documents écrits en \LaTeX . Son utilisation ne nécessite pas de modification notable au source \LaTeX .

Un autre logiciel parfois utilisé est HEVEA. Il a été mis au point par l'INRIA qui le distribue gratuitement. Il est décrit à l'adresse suivante : <http://pauillac.inria.fr/~maranget/hevea/>. Il est supporté sous Windows, Linux (Red Hat et Debian) et Solaris mais utilise les fontes disponibles en standard sous Windows pour fabriquer un document au format HTML, ce qui implique que certaines expressions mathématiques ne sont pas toujours bien rendues.

9 Conclusion

Ce stage a été pour moi une expérience particulièrement enrichissante. En effet, il m'a permis de savoir comment gérer et faire face aux demandes spécifiques de non-informaticiens dans un cadre professionnel, ce qui pour moi est une nouveauté dans la mesure où les précédents stages que j'avais effectués s'étaient déroulés dans des équipes composées exclusivement d'informaticiens. D'autre part il m'a offert l'occasion de me familiariser avec l'écriture de classes \LaTeX , et aussi d'aborder les différents outils existants dans le cadre de l'utilisation et du développement autour de \LaTeX .

J'espère que les résultats de mon travail et de celui de Laurent Candillier pourront servir au moins de base au développement par améliorations successives d'une classe finale d'écriture de thèses françaises à l'Université de Lille1. C'est selon moi un objectif important car à l'heure actuelle le projet GriseMine est l'un des projets de diffusion de littérature grise les plus en pointe au plan national (c'est l'une des 3 actions citées sur le site du ministère). Une forte initiative locale et une forte implication de la communauté des scientifiques de l'université pourraient amener à un rapide développement et ainsi faire de ce projet un projet de référence.