



ÉCOLE DES MINES  
DE DOUAI



OPTION ISIC

INGÉNIERIE DES SYSTÈMES  
D'INFORMATION ET DE  
COMMUNICATION

# PROJET DE FIN D'ÉTUDES

## MandrakeSoft (Paris)

*Aurélien Bompard* (promo 2003)





# Table des matières

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Généralités . . . . .	5
1.2	Avertissement . . . . .	6
1.3	Remerciements . . . . .	6
1.4	Emploi du temps et résumé des tâches effectuées . . . . .	6
<b>2</b>	<b>MandrakeSoft et son domaine d'activité</b>	<b>9</b>
2.1	Le Logiciel Libre . . . . .	9
2.1.1	Un bref rappel : le projet GNU . . . . .	9
2.1.2	L'avènement du noyau Linux . . . . .	10
2.1.3	Les forces du système d'exploitation Linux . . . . .	11
2.1.4	Le Phénomène Linux . . . . .	12
2.1.5	Qui utilise Linux ? . . . . .	13
2.2	Présentation de MandrakeSoft . . . . .	14
2.2.1	Historique de la société . . . . .	14
2.2.2	Un développement et une réussite rapides . . . . .	15
2.2.3	Situation financière de MandrakeSoft . . . . .	16

2.3	MandrakeClub . . . . .	19
<b>3</b>	<b>Travaux réalisés</b>	<b>23</b>
3.1	Présentation de l'existant . . . . .	23
3.2	La fonction de recherche catégorisée . . . . .	24
3.2.1	Modèle de données . . . . .	25
3.2.2	Architecture de la fonction de recherche . . . . .	26
3.2.3	Exemple d'une recherche . . . . .	32
3.2.4	Les scripts d'approvisionnement des tables de recherche . . . . .	33
3.3	La gestion des commentaires . . . . .	35
3.3.1	Description générale . . . . .	35
3.3.2	Description dans le détail . . . . .	37
3.3.3	Conclusion . . . . .	39
3.4	Autres travaux . . . . .	40
3.4.1	Portage de mandrakesoft.com de PHP3 à PHP4 . . . . .	40
3.4.2	Installation du moteur de recherche <i>RpmFind</i> . . . . .	40
3.4.3	Mise en place d'une interface web à l'annuaire LDAP de MandrakeSoft	41
<b>4</b>	<b>Impressions personnelles</b>	<b>43</b>
<b>5</b>	<b>Conclusion</b>	<b>45</b>

## 1.1 Généralités

Le Projet de Fin d'Études est le stage de dernière année à l'École des Mines de Douai, et doit permettre à l'élève de se préparer à la vie professionnelle en le faisant travailler en entreprise sur un projet important.

J'ai choisi de faire mon projet de fin d'études au sein de la société MandrakeSoft, éditeur de la distribution MandrakeLinux. MandrakeLinux fait partie des quatre distributions de Linux les plus utilisées dans monde. J'ai fait ce choix parce que je souhaite travailler ensuite dans le monde du Logiciel Libre, et que MandrakeSoft est un grand nom dans ce domaine. Travailler dans cette société me préparera parfaitement aux spécificités du monde du Logiciel Libre (qui sont nombreuses).

MandrakeSoft traverse actuellement une période difficile. La société est actuellement en redressement judiciaire, depuis janvier dernier, et jusqu'à fin juillet au moins. Cela signifie que la société est protégée par la justice et peut continuer son activité sans devoir régler ses dettes, et en contrepartie elle a le devoir de se réorganiser pour rendre son activité viable. Pendant cette période de six mois, un administrateur judiciaire la surveille, et donnera son verdict à la fin de la période d'observation. Si la société est jugée viable, elle continue son activité et ré-échelonne ses dettes avec ses fournisseurs. Si elle n'est pas jugée viable, c'est la liquidation judiciaire, c'est à dire la fermeture de l'entreprise.

C'est dans ce cadre difficile que j'ai décidé de faire mon projet de fin d'études au sein de cette société, car je pense que c'est une entreprise qui édite un très bon produit, qui a une large base d'utilisateurs, et une bonne réputation. Les causes de sa situation actuelle sont selon moi liées aux décisions de ses dirigeants, et aucunement à un secteur d'activité qui ne serait pas économiquement viable. RedHat, le principal concurrent de MandrakeSoft, l'a prouvé récemment en enregistrant des bénéfices.

Le sujet de mon projet était le développement de ce qui doit devenir le principal moteur de revenus de la société : MandrakeClub. Le Club est un abonnement payant à un certain

nombre de services, que je décrirai plus en profondeur dans les pages suivantes (principalement à la section 2.3 page 19). Un sujet donc très intéressant et très important pour la société. Les revenus liés à ce Club sont en relation directe avec les améliorations apportées au site mandrakeclub.com. Il faut donc constamment y apporter des améliorations visibles et de qualité si on veut attirer des nouveaux membres au Club. Il en résulte que mon projet de fin d'étude sera très important, et participera peut-être au redressement de la société.

## 1.2 Avertissement

On devrait en toute rigueur parler de “GNU/Linux” pour désigner le système d'exploitation à base de noyau Linux que les utilisateurs de MandrakeLinux et de toutes les autres distributions utilisent. J'expliquerai pourquoi en détail dans la section 2.1 page 9. Cependant, dans le reste de ce document, j'appellerai simplement ce système “Linux” parce que c'est sous ce nom qu'il est le plus connu.

## 1.3 Remerciements

Je remercie tout d'abord mon parrain de stage, Denis Havlik, pour les très nombreux conseils, explications, et informations qu'il m'a fourni tout au long de ces quatre mois. J'aimerais remercier aussi Nicolas Chipaux et Charles Davant pour leurs conseils, et Alexis Younes pour toute l'aide qu'il m'a apportée dans mes relations avec les clients. Enfin, j'aimerais remercier Guillaume Delabre sans qui je n'aurais pas eu l'occasion de réaliser ce projet.

## 1.4 Emploi du temps et résumé des tâches effectuées

Les horaires de travail à MandrakeSoft sont très souples. Il est plus important d'avoir fait le travail en temps et en heure que de le faire exactement aux horaires classiques. Cela permet à l'employé de s'organiser au mieux et d'être le plus efficace. Mes horaires de travail étaient en général de 10 heures à 18 heures, avec une pause d'une heure environ pour le déjeuner. La semaine de travail commence classiquement le lundi et se termine

le vendredi. J'ai travaillé environ trois mois sur la création de la fonction de recherche (section 3.2 page 24) et sur la gestion des commentaires (section 3.3 page 35), et ensuite j'ai travaillé à sur différents projets que j'ai regroupé dans la section "Autres travaux" (section 3.4 page 40).





# MANDRAKE<sup>®</sup>SOFT ET SON DOMAINE D'ACTIVITÉ 2

## 2.1 Le Logiciel Libre

### 2.1.1 Un bref rappel : le projet GNU

Pour comprendre Linux et les offres de MandrakeSoft, il convient de revenir quelques années en arrière et d'expliquer les fondements du concept du Logiciel Libre.

Également connu sous le nom générique de Free Software ou d'Open Source, le logiciel libre est né avec l'avènement d'Unix dans les années 1970. Le système d'exploitation est alors livré avec son code source afin que les utilisateurs – en grande majorité des universités et des laboratoires de recherche – puissent en assurer la maintenance. Pour faire face à cette absence d'assistance technique, des communautés d'informaticiens se constituent alors, avec pour objectif le maintien, l'amélioration et la diffusion du système d'exploitation. En 1983, Richard Stallman, chercheur au laboratoire d'intelligence artificielle du MIT, observe la dérive de cet esprit d'entraide vers un modèle fermé et propriétaire qu'il ne veut pas accepter. Il fonde la Free Software Foundation, une organisation militant pour l'emploi des logiciels libres, et démissionne du MIT pour que l'université ne puisse pas revendiquer son travail. Son projet est de créer un système complet entièrement libre, baptisé GNU (Gnu is Not Unix, acronyme récuratif soulignant la similarité de GNU avec Unix). Pour protéger la liberté des logiciels du projet GNU, Stallman se fait aider par des juristes et écrit la GPL, General Public License, qui définit les conditions d'utilisation et de distribution du logiciel libre. Stallman décrit le logiciel libre comme respectant les quatre libertés fondamentales de l'utilisateur :

- La liberté d'exécuter le programme, pour tous les usages (liberté n°0).
- La liberté d'étudier le fonctionnement du programme, et de l'adapter à vos besoins (liberté n°1). Pour ceci l'accès au code source est une condition requise.
- La liberté de redistribuer des copies, donc d'aider votre voisin, (liberté n°2).

- La liberté d'améliorer le programme et de publier vos améliorations, pour en faire profiter toute la communauté (liberté n°3). Pour ceci l'accès au code source est une condition requise.

En résumé, un utilisateur est libre de faire tout ce qu'il veut d'un logiciel libre, y compris de le modifier. Mais s'il décide de le modifier *et de le redistribuer*, alors les modifications qu'il y a apportées doivent être publiées sous la licence GPL.

### 2.1.2 L'avènement du noyau Linux

Le système GNU se développe, mais au début des années 90, il manque encore un composant essentiel : le noyau. C'est le coeur du système d'exploitation, qui fait le lien entre le matériel et le logiciel. Le système GNU devait donc alors encore être installé sur un noyau propriétaire, ce qui lui faisait perdre beaucoup de son intérêt.

C'est alors qu'un étudiant finlandais, Linus Torvalds, qui travaillait sur *Minix*, un autre système d'exploitation cloné d'UNIX, décide d'écrire un noyau qui fonctionnerait sur son 386. Il place son travail sous Licence GPL, et annonce sur un forum dédié qu'il a commencé à travailler là dessus et qu'il accepte avec joie toutes les contributions. Le texte original peut être retrouvé ici : <http://www.linux10.org/history> (on y remarquera même la phrase suivante : «[Linux] n'est qu'un passe-temps, ce ne sera pas aussi important et professionnel que GNU.» ) L'accueil de son annonce va au-delà de toutes ses espérances, et Linux commence à se développer de façon fulgurante.

Associé au noyau Linux, le projet GNU offre donc un système d'exploitation utilisable et complètement libre au sens de la GPL. C'est pour cela que l'on devrait parler de GNU/Linux, mais j'utiliserai le nom Linux pour désigner le système d'exploitation dans la suite de ce document, et préciserai qu'il s'agit du noyau lorsque ce sera nécessaire.

Si le terme "libre" implique que le code source accompagne le logiciel, il ne signifie en aucun cas que l'application est forcément gratuite. La licence GPL empêche simplement un distributeur de s'emparer du travail d'une communauté de développeurs et d'y inclure des restrictions d'utilisation. Aujourd'hui, la licence GPL régit l'immense majorité des programmes fonctionnant sous Linux.

Il faut faire la différence entre le mouvement du Logiciel Libre initié par Richard Stallman au milieu des années 80, et le mouvement Open Source, qui commença vers la fin

des années 90. Ce mouvement prône le partage du code source non pas pour des raisons éthiques, mais parce que c'est un mode de développement plus efficace, qui permet de créer des applications plus stables et plus sécurisées. Les moyens sont les mêmes mais la fin est différente.

Sous les feux de l'actualité et présenté comme une alternative à Windows de Microsoft©, le système d'exploitation Linux est le produit "abouti" d'un noyau Unix conçu en 1991. Depuis 1994, date de la diffusion de la version 1.0 de Linux sur Internet, le noyau a été enrichi et optimisé grâce à un développement collaboratif. Il se révèle aujourd'hui comme l'un des systèmes les plus performants et les plus stables du marché.

### 2.1.3 Les forces du système d'exploitation Linux

Une des grandes forces du système d'exploitation Linux réside dans la nature même de son développement. Le développement des applications Open Source via Internet repose aussi bien sur des équipes internes à une entreprise que sur des communautés virtuelles indépendantes constituées d'experts en informatique et d'utilisateurs. La croissance du média Internet a favorisé très significativement la participation en nombre des contributeurs externes et ainsi amélioré de manière spectaculaire la rapidité du développement et la qualité des produits Open Source. L'accès au code source du système d'exploitation Linux permet aux utilisateurs expérimentés de personnaliser et d'optimiser le système en fonction de leurs besoins spécifiques. Il possède la flexibilité nécessaire à son adaptation dans un grand nombre d'environnements

De plus, il est compatible avec toutes sortes de matériels, même obsolètes. Il contribue ainsi à la pérennité des investissements aussi bien des entreprises que des particuliers : ceux-ci n'ont plus à mettre au rebut leurs "vieux" ordinateurs. Bien entendu, Linux donne sa pleine mesure avec des équipements récents où les applications classiques (Oracle par exemple) portées sous Linux s'avèrent de 20 à 30% plus rapides. Autre avantage, Linux est disponible sur les principales plates-formes du marché : PowerPC d'Apple, i386, Pentium I, II et III d'Intel et compatibles, Ultra Sparc de Sun, Alpha de Digital. Par exemple, la distribution *Debian GNU/Linux* supporte onze architectures différentes. En d'autres termes, Linux est conçu pour fédérer et faire communiquer les plates-formes d'un parc informatique hétérogène. Enfin, comme Linux autorise la personnalisation du code source des programmes selon leurs besoins, de nombreuses entreprises, qui ont investi massivement

dans des systèmes propriétaires, voient aujourd'hui sa grande pertinence. En 1999, Linux s'associait à de grands acteurs du marché tels qu'IBM, Hewlett-Packard, Compaq, Dell, Oracle, Sun, SGI, Netscape, Apple.

### 2.1.4 Le Phénomène Linux

Même s'il est difficile de se baser sur des statistiques fiables concernant le nombre d'utilisateurs, les analystes d'IDC, de Dataquest et du Gartner Group estiment que le nombre d'utilisateurs se situe entre 9 et 15 millions (ventes des distributions Linux). Si les téléchargements via Internet étaient pris en compte (logiciel libre), le nombre réel des utilisateurs du système d'exploitation Linux serait de l'ordre de 20 millions. Ces statistiques datent de 2001.

Le marché Serveurs Linux est devenu en quelques années un système majeur sur le marché des serveurs. Il se positionne comme l'alternative de Windows NT de Microsoft, de SCO Unix et de Novell Netware. En 1999, Linux représentait 25% des nouvelles installations de systèmes d'exploitation serveurs : 1,35 millions de solutions Linux ont été installées sur un marché global de 5,69 millions (Source IDC). Linux se place ainsi en deuxième position juste derrière Windows NT mais devant Netware et Unix. Selon Dan Kusnetzky, analyste chez IDC, « Linux se développe plus vite que nous le pensions. Nous avons estimé que Linux serait N° 2 en 2002 ou 2003. C'est arrivé dès 1999. » La croissance des installations de solutions Linux, 93% de 1998 à 1999, s'accélère et est nettement supérieure à la croissance moyenne pour l'ensemble des systèmes d'exploitation, de l'ordre de 23%. Ce succès tient à deux facteurs essentiels : le ralliement au phénomène Linux des grands noms des éditeurs de logiciels (Oracle, Corel, Netscape...) et des constructeurs informatiques (IBM, Compaq, Dell, Hewlett-Packard...) l'explosion de la demande pour les serveurs web au sein desquels la pénétration de Linux est très forte. Selon IDC, en 2004 de tous les systèmes d'exploitation pour serveurs, seuls Windows NT/2000 et Linux connaîtront une croissance et des parts de marché significatives. En 2004, Linux représentera 30,3% du marché clients/serveurs en base installée. Cette même année, Windows NT/2000 (44,6%) et Linux (38,2%) représenteront à eux deux les 2 tiers des nouvelles installations de systèmes d'exploitation. Entre 1999 et 2004 Linux générera le taux de croissance annuel pondéré le plus important tant en termes de ventes de nouvelles licences (28,4%) qu'en termes de revenus (23,3%) (Source IDC).

### 2.1.5 Qui utilise Linux ?

Qui sont les utilisateurs de Linux ? Linux est le seul système d'exploitation Unix dont la part de marché augmente chaque année. Ses utilisateurs se répartissent en différentes catégories : La première catégorie se compose d'utilisateurs inconditionnels qui sont, en règle générale, des connaisseurs de l'informatique, capables d'apprécier la stabilité et les performances du système à sa juste valeur.

De plus en plus de néophytes s'intéressent à une solution complète (système d'exploitation et applications logicielles) qui leur assure la performance, la stabilité, la simplicité et la convivialité à un coût très abordable. L'essor et la multiplication de supports presse généraliste dédiée à Linux attestent de l'intérêt et de l'engouement du grand public pour Linux.

Les PME/PMI : outre l'aspect économique non négligeable du logiciel libre (frais de licences réduits), Linux fonctionne sur n'importe quelle plate-forme, même techniquement dépassée. De plus, la fiabilité du système ne requiert pas la présence, au quotidien, d'un ingénieur système, alors qu'elle est indispensable dès que l'on opte pour Netware, SCO Unix et Windows NT.

Les grands comptes, tels que les opérateurs télécom et les fournisseurs d'accès Internet (le serveur Web Apache représente 50% des serveurs des réseaux). La fiabilité et l'efficacité du système d'exploitation remportent l'adhésion de ces clients exigeants. Les universités et centres de recherche bénéficient grâce à Linux non seulement de plates-formes de développement et d'enseignement-formation, mais également de cluster de calcul scientifique et de bureautique scientifique.

Enfin, les industriels reconnaissent les qualités de Linux et l'intègrent à leur offre produit. La possibilité d'accéder au code source du produit par la GPL fait de Linux le choix N°1 pour les logiciels embarqués ou intégrés à tout type de produit (PABX, moniteur, boîtier TV, téléphone mobile, organisateur personnel...)

L'objectif de MandrakeSoft est la démocratisation de Linux.

### 2.2 Présentation de MandrakeSoft

MandrakeSoft est un distributeur du système d'exploitation GNU/Linux, développé par des volontaires dans le monde entier. Son objectif est de faire bénéficier les utilisateurs des innovations issues des travaux de la communauté, qui réunit plusieurs milliers de personnes à travers le monde, en proposant un système d'exploitation puissant, stable associé à une gamme complète de solutions pour les postes clients (environnement graphique rappelant MS Windows, suite bureautique compatible avec tout format de fichier, navigateur Internet) et pour les serveurs (serveur web, serveur de messagerie, serveur de fichiers et d'impression...) soit environ 2300 applications à ce jour.

Aujourd'hui, Mandrake Linux est la distribution Linux la plus internationalisée puisqu'elle est supportée dans près de 50 langues à l'installation. De plus, Linux-Mandrake garantit une installation et une prise en main simples, rapides et fiables. Les différentes solutions Mandrake Linux regroupent le système d'exploitation, les programmes associés sur des CD-ROM, les outils d'installation, une documentation complète ainsi que du support technique. MandrakeSoft va plus loin en proposant à ses clients un accompagnement global en termes de services sous forme de solutions packagées et/ou de solutions sur mesure.

#### 2.2.1 Historique de la société

MandrakeSoft est créée en 1998 par Gaël Duval, Jacques Le Marois et Frédéric Bastok dans le but de démocratiser le Logiciel Libre. Passionnés par ce mouvement encore assez récent, les fondateurs décident d'adopter la solution technique suivante : La distribution MandrakeLinux sera fondée sur la distribution RedHat, mais en y ajoutant l'environnement graphique KDE. KDE ne pouvait pas être intégré à RedHat parce qu'à l'époque, la librairie graphique QT sur laquelle il était fondé n'était pas sous licence GPL. C'est ce problème de licence qui a poussé les développeurs du libre à créer l'environnement concurrent : GNOME. Mais à l'époque (et encore aujourd'hui), l'environnement KDE lui était supérieur à de nombreux points de vue. Les fondateurs de MandrakeSoft ont donc voulu associer la meilleure distribution, RedHat, au meilleur environnement graphique possible. C'était leur recette du succès, qui a bien fonctionné puisque la société a rapidement connu une assez grande popularité.

Au début, MandrakeLinux a suivi RedHat en terme de numéros de version, pour bien

signifier qu'ils offraient une compatibilité à 100% avec la distribution la plus répandue à l'époque, et encore de nos jours. Mais à partir de la version 7.0, MandrakeLinux s'est détaché de RedHat, pour montrer qu'ils étaient devenus plus que la somme de RedHat et de KDE. En effet, la distribution MandrakeLinux est maintenant très différente d'une RedHat, bien qu'elle soit toujours compatible, c'est à dire qu'un package prévu pour RedHat s'installera bien sur une Mandrake (dans une certaine limite bien sûr). L'avantage le plus significatif de MandrakeLinux par rapport à RedHat Linux est le programme **urpmi** qui permet d'installer automatiquement des paquetages logiciels RPM en résolvant les dépendances et même en les téléchargeant à partir d'internet si nécessaire. Aujourd'hui, la distribution comporte environ 3000 paquetages logiciels, et est déclinée en 3 versions :

1. la **Standard** (environ 30€), destinée à une utilisation basique,
2. la **PowerPack** (environ 80€), destinée à ceux qui ont une utilisation plus poussée de leur ordinateur (comme par exemple le développement),
3. la **Prosuite** (environ 150€), destinée aux serveurs d'entreprise.

Il existe aussi une édition **Corporate Server**, qui est destinée aux serveurs critiques des grandes entreprises. C'est une version qui ne sort que tous les deux ans environ, et qui bénéficie d'un support accru de la part de MandrakeSoft. Le rythme de parution des autres versions de la distribution est d'environ une tous les 6 mois. Ce rythme rapide est rendu nécessaire par la vitesse de développement du monde Open Source.

### 2.2.2 Un développement et une réussite rapides

Plusieurs facteurs expliquent la croissance exceptionnelle de l'entreprise, passée de 3 à plus de 120 salariés en 2 ans : Dans le respect de la philosophie du Logiciel Libre, MandrakeSoft met librement à disposition sur le Web ([www.linux-mandrake.com](http://www.linux-mandrake.com)) le code source de sa distribution Linux-Mandrake. Ce sont ainsi plusieurs centaines de développeurs indépendants dans le monde qui améliorent en permanence le produit dans la version de développement appelée Cooker, et en font une des solutions Linux les plus simples, complètes et puissantes. De plus la société est implantée en France et aux États-Unis. L'excellence technologique des produits de MandrakeSoft est reconnue par les spécialistes du marché. Linux-Mandrake a été désignée "Meilleur produit de l'année" et "Meilleure distribution pour serveur" lors du salon LinuxWorld Expo 1999 ; elle a reçu en avril 2000 le "Platinum Award" de la revue britannique PC Answers ainsi que l'Editor Choice Award de

la célèbre revue américaine Linux Magazine en septembre 2000. Numéro 1 en France, Linux-Mandrake s'est également positionnée en première place des ventes de produits Linux aux États-Unis en juillet août et septembre 2000 avec plus de 30% du marché Nord américain (Source : PC Data). La réussite de Mandrakesoft démontre que l'industrie américaine du logiciel peut facilement être concurrencée à partir de la France. MandrakeSoft a été agréée "entreprise innovante" par l'Anvar. Elle est l'une des premières sociétés éditeurs de distributions Linux à avoir reçu cet agrément pour son savoir-faire en logiciel Open Source. Enfin, MandrakeSoft s'implique fortement dans de nombreuses actions ayant pour but la défense et la promotion du Logiciel Libre et de l'Open Source.

Des partenariats "Écoles-Universités" sont en cours, dont celui avec l'ENSTA dans le cadre d'un projet d'enseignement et pour le développement de Logiciels Libres.

### 2.2.3 Situation financière de MandrakeSoft

MandrakeSoft a été créée en 1998, et a été assez rapidement profitable : dès 1999. En février 1999, les premières boîtes sont mises sur le marché, et neuf mois plus tard la société atteint un chiffre d'affaires de 571000€, pour un bénéfice de 55000€.

Les investisseurs commencent à s'intéresser à MandrakeSoft, et apportent 18 millions d'euros. Au même moment, RedHat vaut 23 milliards de Dollars au Nasdaq, et est le leader mondial des solutions Linux.

En avril 2000, MandrakeSoft embauche Henri Poole, fondateur de Vivid Studios (une start-up américaine de services internet). Henri Poole recrute ensuite lui-même 5 collaborateurs directs, tous grassement rémunérés (entre 180 000 et 300 000\$). L'entreprise loue alors de locaux boulevard Sébastopol à Paris, et la nouvelle équipe réoriente la stratégie vers l'e-Learning. Le but est de transformer MandrakeSoft en dotcom et donc d'attirer de nouveaux investisseurs.

Les dépenses augmentent de 400%, et l'entreprise a commencé à accumuler les dettes. Cela a été un désastre financier, que les dirigeants ont découvert au printemps 2001. Entre janvier et mars, MandrakeSoft, qui emploie alors 150 personnes, a dépensé 4,7 millions d'euros pour un chiffre d'affaires inférieur à 400 000€.

Les dirigeants sont remerciés, et MandrakeSoft se recentre vers son activité principale :



l'édition de la distribution MandrakeLinux. Les coûts sont divisés par 4 en quelques mois. En juillet 2001, MandrakeSoft réussit son introduction au Marché libre. Mais la société est très en retard dans son développement : elle commence tout juste à développer la prestation de services alors que RedHat réalise déjà plus de 80% de son chiffre d'affaire sur le conseil, la maintenance et la formation. En 2002, les services représentent moins de 10% des ventes chez MandrakeSoft.

Le graphique 2.1 page suivante montre une progression impressionnante du résultat d'exploitation (en K€). Le graphique 2.2 page suivante montre l'accroissement du chiffre d'affaires.

En décembre 2001 est créé le club d'utilisateurs Mandrake, ou MandrakeClub. C'est un club payant auquel les utilisateurs peuvent souscrire pour obtenir des avantages supplémentaires. Évidemment, à la création du club, les avantages étaient plus ou moins inexistantes. Il consistait en un site web sous phpNuke, un moteur de site web très répandu. En avril 2002, Mandrake est en grave crise financière, et demande officiellement à ses utilisateurs de s'abonner au club pour les aider à traverser cette passe difficile. Cela a apporté une masse importante d'utilisateurs au Club, et une quantité importante de cash dans les caisses de Mandrakesoft. Mais cela ne suffit pas, car en décembre 2002, un deuxième appel du même type est lancé. Là aussi, il rime avec une arrivée en masse d'utilisateurs au club. Malheureusement, les dettes sont trop importantes, et la société dépose le bilan un mois plus tard.

MandrakeSoft est en cessation de paiements depuis le 13 Janvier, et en redressement judiciaire depuis le 27 Janvier pour une période d'observation de 6 mois. Les objectifs sont d'assainir la société et de renégocier les dettes, avec comme perspective la continuation de l'activité avec une équipe réduite, ou la liquidation judiciaire en cas d'échec.

Le défi est grand, car en 2002 la société enregistrait une perte de 6 millions d'euros, pour un chiffre d'affaires de 4,7 millions. MandrakeSoft doit sortir de la période d'observation le 27 Juillet, c'est la date à laquelle le juge rendra son verdict.

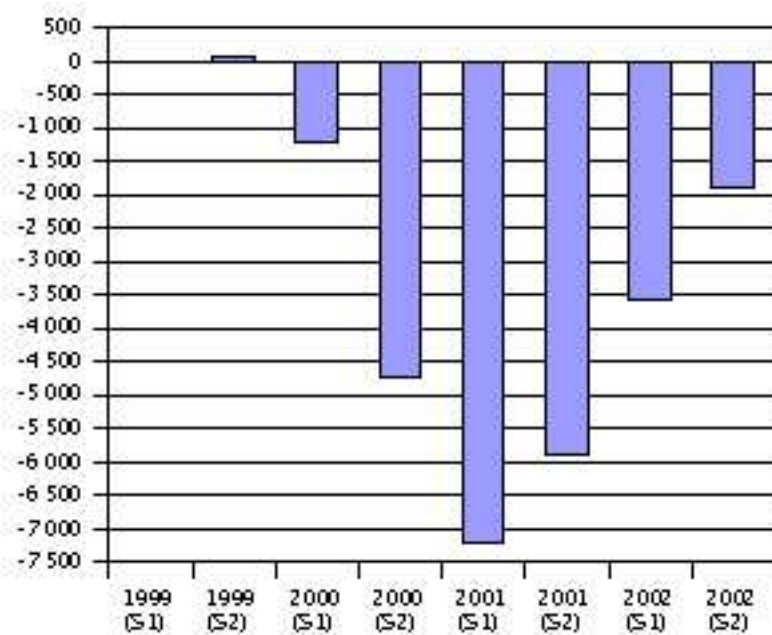


FIG. 2.1 – Résultat d'exploitation (en K€)

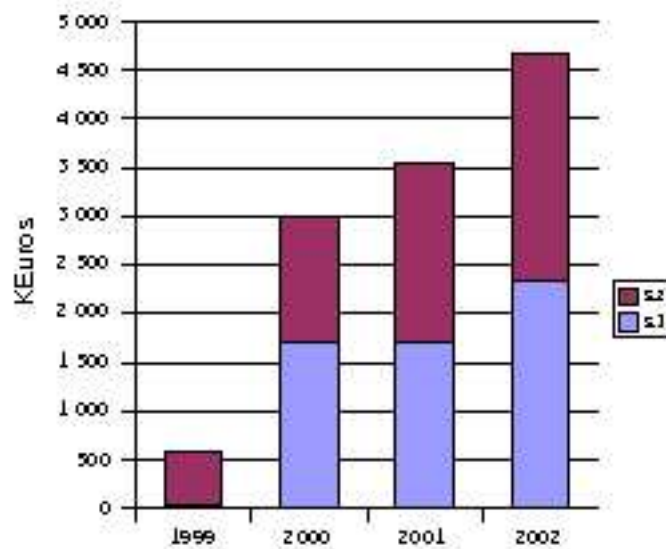


FIG. 2.2 – Chiffre d'affaires (en K€)

### 2.3 MandrakeClub

MandrakeClub est créé en 2001, à l'initiative de mon parrain Denis Havlik, et constitue une manière révolutionnaire d'envisager le financement de la société. Avec la multiplication des accès à internet à haut débit, le nombre de boîtes vendues à des individus va très probablement chuter, au profit des téléchargements gratuits sur internet. Comme la clientèle de MandrakeSoft est constituée principalement de particuliers, ils est important de les attirer vers un modèle financier rémunérateur pour MandrakeSoft, sans quoi la société risque de voir son nombre d'utilisateurs augmenter sans que le chiffre d'affaires suive.

Le concept de MandrakeClub est simple : créer un club payant d'utilisateurs qui auront accès à plus de services de la part de leur distributeur Linux. Ces avantages, peu nombreux à la naissance du club, se sont peu à peu étoffés pour offrir aujourd'hui une source importante d'information et de services. A la création de MandrakeClub, le site web offrait un contact privilégié avec les dirigeants de l'entreprise par le biais d'articles et de commentaires, et donc le sentiment pour les membres du club que leur voix est écoutée. Le site proposait aussi en téléchargement les logiciels commerciaux qui se trouvent dans les éditions packagées de la distribution, et pas dans les images disque offertes au téléchargement.

Par la suite, le site web MandrakeClub s'est beaucoup développé, et il offre aujourd'hui les avantages suivants :

- des articles en plusieurs langues, traduits par des volontaires,
- des forums où on peut discuter de l'avenir de la société et de la distribution, où on peut chercher de l'aide sur un problème technique, ou aider les autres membres du club,
- des logiciels supplémentaires à télécharger. Il ne s'agit pas seulement là des logiciels commerciaux des versions "boîte" de la distribution, mais aussi de n'importe quel logiciel que le membre du club voudrait voir adapté et packagé pour sa distribution MandrakeLinux. Le membre dispose d'un système de vote pour annoncer les programmes qu'il aimerait obtenir, et ensuite une équipe de volontaires réalise les packages et les propose aux autres membres.
- le site permet aussi aux membre de faire savoir à MandrakeSoft quels packages ils aimeraient voir dans la prochaine version de la distribution, et ce en utilisant le système de vote précédemment décrit.
- enfin, mandrakeclub.com propose à la consultation ou au téléchargement tout un

ensemble de manuels et de documentation en plusieurs langues.

Tous ces avantages font de MandrakeClub un point central pour l'utilisateur de MandrakeLinux.

Le club compte à ce jour entre 15000 et 20000 membres, et représente à peu près un tiers des revenus de MandrakeSoft.

La figure 2.3 montre l'évolution des revenus générés par le club en milliers de Dollars au travers des derniers mois. Au mois de juillet, le club s'est vu ajouter enfin quelques

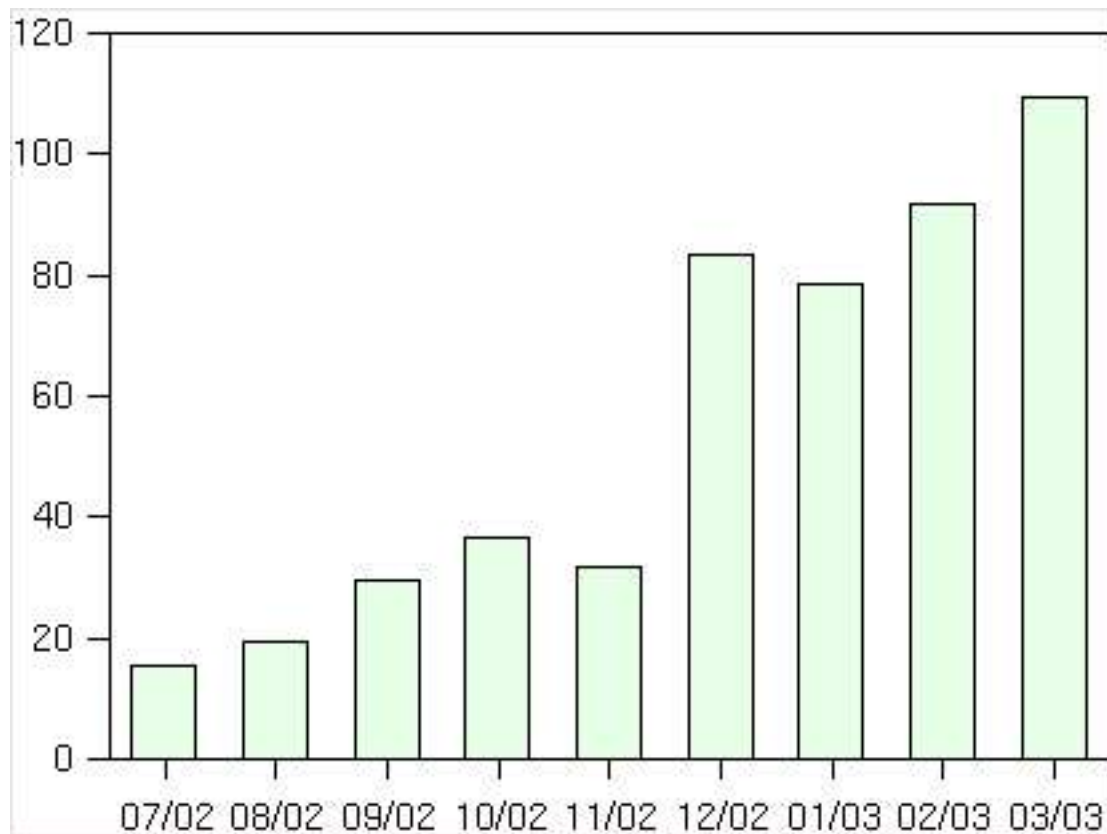


FIG. 2.3 – Revenus générés par MandrakeSoft, en K\$

nouveaux services, donc les revenus ont augmenté. En décembre, MandrakeSoft a lancé un appel à l'aide à ses utilisateurs, ce qui explique le doublement des revenus du club pour décembre 2002. En janvier, nous avons lancé le service de "vote pour les RPM de la 9.1", qui a permis aux membres de choisir les packages qu'ils veulent voir dans la prochaine version de la distribution. Cette initiative a été fortement couverte par la presse, et aussi fortement appréciée. Et au mois de mars 2003 a été lancée la nouvelle version de la distribution

MandrakeLinux. Les acheteurs de la distribution reçoivent maintenant automatiquement un mois d'abonnement gratuit. Ce système augmente les revenus car beaucoup de personnes veulent conserver leurs avantages après la fin de la période d'essai gratuite.

Le MandrakeClub est donc une source très importante de revenus de la société, bien que très jeune. MandrakeSoft s'oriente maintenant vers une plus grande mise en avant du Club, et compte de plus en plus dessus, notamment grâce aux marges extraordinaires qu'apportent le service. A terme, le Club doit devenir un point central du fonctionnement de la société. Si MandrakeSoft survit à cette période difficile, et si MandrakeSoft est encore en activité aujourd'hui, c'est en grande partie grâce à MandrakeClub.



### 3.1 Présentation de l'existant

Mon travail au sein de MandrakeSoft est d'assister mon parrain Denis Havlik, chef de projet MandrakeClub, dans l'amélioration et la gestion du site web.

Lorsque la création du Club a été décidée, MandrakeSoft avait besoin de mettre en place un site web rapidement pour pouvoir rapidement commencer à recevoir l'argent des cotisations. Denis Havlik a alors choisi un système de gestion de contenu très répandu sur internet à ce moment là : phpNuke. Ce moteur est aussi la base de plusieurs autres sites web de la société. phpNuke a l'énorme avantage de proposer un système prêt à l'emploi, fonctionnel, et proposant beaucoup de fonctionnalités. phpNuke est sous licence GPL, mais malheureusement il a deux gros défauts :

1. l'auteur n'est pas du tout intéressé par les principes de l'Open Source,
2. le code PHP est de très mauvaise qualité.

La conséquence du premier point est qu'il est inutile de proposer à l'auteur d'inclure dans le code principal les améliorations que l'on y a apporté. Denis Havlik avait déjà essayé, mais l'auteur, F. Burzi, a simplement récupéré le code et ajouté son propre copyright dessus. De plus, F. Burzi est en train de réfléchir à une réécriture de phpNuke à partir de zéro, sous une licence propriétaire, en cachant les sources grâce à l'encodeur Zend. Travailler avec phpNuke est donc assez risqué, car il y a des chances pour que ce moteur de site web n'ait pas d'avenir dans le monde Open Source.

Les conséquences du second point sont nombreuses. Francisco Burzi a déclaré avoir appris le langage PHP en écrivant phpNuke, et cela se voit. Tout d'abord, phpNuke contient de nombreuses failles de sécurité, et est très souvent l'objet d'alertes de sécurité. Nous avons eu à faire face à une faille de sécurité grave environ deux mois après le début du stage.

Ensuite, le code est très difficile à maintenir et à faire évoluer. La totalité du système de gestion de contenu phpNuke contient environ une dizaine de commentaires explicatifs, et les noms des variables sont parfois aussi clairs que `$something`. De plus, le système contient

énormément de duplication de code, ce qui entraîne qu'une modification doit souvent être répercutée dans plusieurs fichiers, faisant chacun autour de 900 lignes.

Vous trouverez plus d'informations sur le système de gestion de contenu phpNuke sur le site du projet : <http://www.phpnuke.org>.

L'année dernière, un étudiant en stage avait apporté des améliorations importantes au site web :

- Site multilingue : tous les articles pouvaient être traduits dans n'importe quelle langue par des volontaires, et les membres du Club peuvent définir les langues qu'ils comprennent dans leur préférences.
- Vote RPM : les membres du Club peuvent voter pour les logiciels qu'ils préfèrent, et des volontaires peuvent ensuite les packager pour la distribution MandrakeLinux.

Ces améliorations rendent impossible aujourd'hui la migration vers un autre système de gestion de contenu, tel que PostNuke, GeekLog, ou SPIP.

## 3.2 La fonction de recherche catégorisée

Le site web <http://www.mandrakeclub.com> contient beaucoup d'informations. Parmi les plus importantes sont :

- les articles
- la documentation
- les commentaires sur les articles et la documentation
- les forums
- les paquetages logiciels RPM.

Le site doit devenir une importante source d'information pour les membres, il est donc nécessaire de l'équiper d'une bonne fonction de recherche. Comme les données contenues sur le site peuvent être de nature très différentes, nous avons décidé de développer une fonction de recherche qui classerait les informations par catégorie et sous-catégorie, et pourrait ensuite rechercher exclusivement dans ces catégories et sous-catégories. Ainsi, l'utilisateur pourra faire une recherche parmi les articles, la documentation, les forums, les RPMs, et les commentaires, mais aussi parmi les articles et les commentaires ensemble, ou encore seulement un certain manuel et non pas la totalité de la documentation.



### 3.2.1 Modèle de données

Nous avons opté pour un jeu de tables supplémentaires dans la base de données, où seraient stockées les informations à rechercher. Ces tables sont alimentées par des scripts qui extraient les informations intéressantes de la base de données, qui les traitent, et les rechargent dans les tables de recherche. Le modèle de données est décrit par le schéma en figure 3.1.

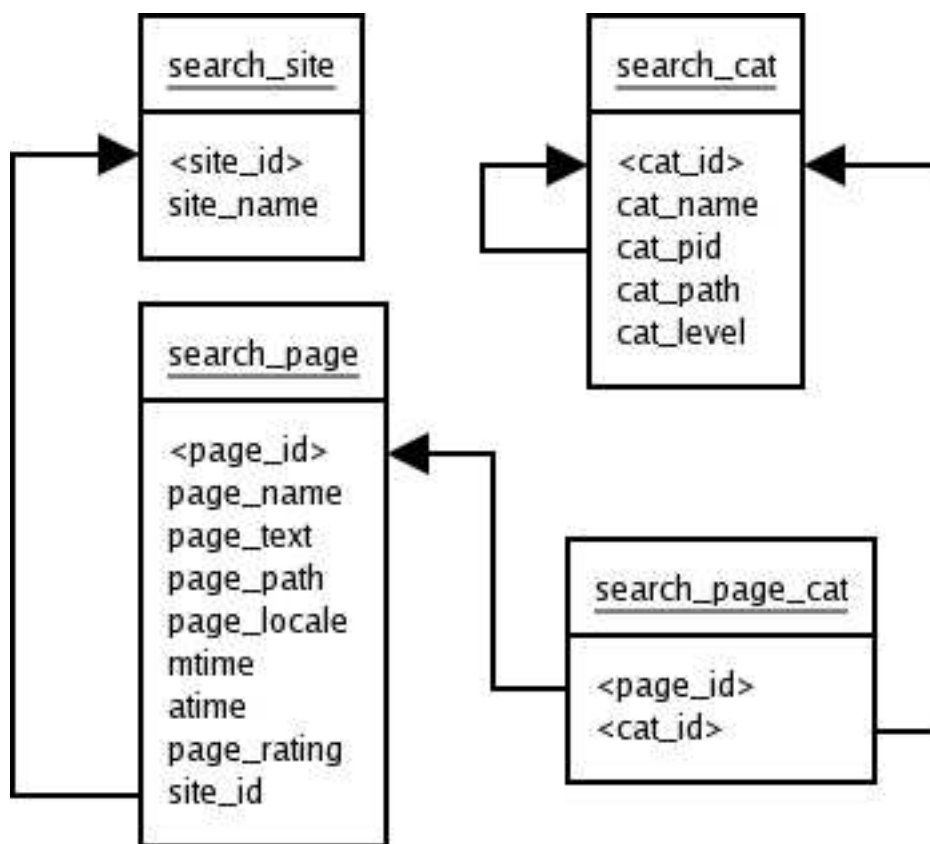


FIG. 3.1 – Modèle de données de la fonction de recherche

**La table search\_cat** Le champ cat\_path est du type “/article.php ?sid=%” pour pouvoir associer une catégorie à un modèle d’URL. L’intérêt de ce champ sera explicité par la description de la fonction ParsePath() page 27. On notera qu’une page (c’est à dire un commentaire, un manuel, un article. . .) peut être associée à plusieurs catégories. Le champ cat\_level est utilisé pour créer une arborescence des catégories. Dans cette arborescence, les catégories qui sont au même niveau ont le même champ cat\_level. Par exemple, un manuel

peut décrire une certaine version de la distribution, et ce pour une certaine architecture : i586 (PC) ou PPC (Macintosh). Ces deux catégories auront donc le même champ `cat_level`. Par contre, un certain manuel dépend de la catégorie générale “Documentation”, et aura donc un champ `cat_level` plus élevé.

**La table `search_page`** Le champ `page_name` correspond au titre de la page, `page_text` correspond au texte qui est indexé, `page_path` est l’URL de la page, et `page_locale` est la langue dans laquelle est écrite le texte. Les champs `page_mtime` et `page_atime` sont utilisés pour supprimer les pages trop vieilles. Le premier correspond à la date de création de la page, et le deuxième est mis à jour à chaque fois que le script de mise à jour trouve cette page dans la base de données du site. Ainsi, il est facile de chercher les pages dont le champ `page_atime` dépasse une semaine, et on trouvera les pages qui ont disparu du site depuis environ une semaine. On peut alors les supprimer de la base de recherche. Enfin, `page_rating` n’est pas encore utilisé, mais pourrait servir à donner une note à la page et donc à améliorer les résultats fournis à l’utilisateur.

La signification des autres tables est claire. La recherche des mots dans la base est effectuée par un index FULLTEXT dans les champs (`page_name`, `page_text`).

#### 3.2.2 Architecture de la fonction de recherche

Pour la fonction de recherche en PHP, nous avons choisi une architecture basée sur une classe `search` qui est instanciée lorsqu’une recherche est demandée. Cette classe contient les variables d’instance suivantes :

- `$dbi`
- `$search_text`
- `$cats`
- `$search_locale`
- `$page_id`

**`$dbi`** C’est la ressource de la base de données. Cela correspond à la connexion courante au serveur MySQL. Cette variable est utilisée pour réutiliser la même connexion pour faire plusieurs recherches.

**\$search\_text** C'est le texte à rechercher. Il est fourni par le formulaire de recherche si on vient d'une autre page du site que la page de recherche. Si on arrive directement à la page de recherche sans passer par un champ de recherche ailleurs sur le site, alors ce champ est vide. Dans ce cas précis, la page de recherche n'affichera pas la zone de résultats.

**\$cats** Cette variable correspond à une liste (array) des catégories sélectionnée par l'utilisateur. Cette liste est extraite de l'URL passée en paramètre au constructeur (voir plus loin page 32), mais peut aussi être passée en tant que champ de type "hidden" par le formulaire de recherche situé sur toutes les pages.

**\$search\_locale** C'est la ou les langue(s) dans laquelle (lesquelles) l'utilisateur veut les résultats. La variable est elle aussi déterminée à partir de l'URL passée en paramètres, et peut elle aussi être fournie par le champ de recherche. La liste des langues se voit automatiquement ajouter les langues des préférences de l'utilisateur si elles ne sont pas déjà présentes.

**\$page\_id** Cette variable est l'adresse de la page passée en paramètre au constructeur (voir plus loin page 32).

La classe contient aussi les méthodes suivantes :

- ParsePath()
- DisplayForm()
- ShowMenus()
- FormatResult()
- ShowResults()
- GenPage()
- Search()

**ParsePath()** Cette fonction essaye de déterminer la catégorie dans laquelle se trouvait l'utilisateur en utilisant la fonction "LIKE" de MySQL, comme on peut le voir dans l'exemple de code ci-dessous :

```
$query="SELECT cat_id FROM search_cat where '"
```

```
. addslashes($this->page_id)
.'" LIKE cat_path";
```

**DisplayForm()** Si le paramètre “simple” est passé à cette fonction, elle génère le champ de recherche. Sinon, elle affiche en plus les menus, en utilisant la fonction ShowMenus().

**ShowMenus()** Cette fonction génère les menus déroulants en lisant la liste des catégories sélectionnées, et en positionnant les champs correspondants suivant la valeur de leur “level” (voir la description de la table search\_cat page 25). Par exemple, si on a choisi la catégorie “Documentation”, le menu déroulant suivant proposera la liste des manuels disponibles. Si on a choisi la catégorie “Forums”, le menu déroulant suivant proposera la liste des forums du site. Cette fonction est conçue pour pouvoir être appelée récursivement. L’exemple de code ci-dessous montre comment ces menus déroulants sont générés.

```
function ShowMenues($cat_pid=0) {
    global $dbi;
    // $cat_pid is not user-provided variable, but...
    if (!is_numeric($cat_pid)) $cat_pid=0;
    static $cat_level_old=0;
    if ($cat_pid==0)
        $ret .= "<SELECT name=\"cat_ids[]\"
                onChange=\"submit()\"><OPTION>All</OPTION>";
```

On voit ici que tous les menus déroulants comportent une option “All” pour sélectionner tous les choix disponibles. Par exemple, pour chercher dans tous les forums, ou dans toute la documentation.

```
$q="SELECT cat_id, cat_name, cat_level, cat_path
    FROM search_cat
    WHERE cat_pid='$cat_pid' ORDER BY cat_level";
```

Ici, on récupère toutes les catégories qui sont dépendantes de la catégorie “\$cat\_pid” et on les classe par “\$cat\_level”. Si la fonction est appelée sans paramètre, alors la variable “\$cat\_pid” vaut 0, et on récupère les catégories qui n’ont pas de parent, c’est à dire les catégories de niveau le plus élevé. Elles sont pour l’instant au nombre de quatre : Articles, Documentation, Forums et Commentaires.

```
$result=sql_query($q, $dbi);
while ( list( [...] ) = sql_fetch_row($result, $dbi)) {
    if ($cat_level != $cat_level_old) {
        // If the level changed...
        $cat_level_old=$cat_level;
        if(!empty($match_pid)) { // If we found a match
            $ret .= $this->ShowMenues($match_pid);
            $match_pid=0;
        }
        $ret.="</SELECT>\n&nbsp; : &nbsp;
            <SELECT name=\"cat_ids[]\"
            onChange=\"submit()\"><OPTION>All</OPTION>";
    }
}
```

On parcourt toutes les catégories données par la recherche ci-dessus. Si le niveau change et qu’une catégorie fait partie de la liste des catégories choisies, on rappelle la fonction récursivement pour générer les menus déroulants qui dépendent de cette catégorie. Si le niveau a changé mais que la catégorie courante ne fait pas partie de la liste des catégories choisies, on génère un menu déroulant (balise *SELECT*).

```
if (in_array($cat_id, $this->cat_ids)) {
    $match_pid=$cat_id;
    $sel="selected";
}
$ret.="<OPTION value=\"$cat_id\" $sel>$cat_name</OPTION>";
$sel="";
}
```

Si la catégorie fait partie de la liste, on prépare la mention “selected” au menu déroulant, qui fixe l’option par défaut. Dans tous les cas, on génère le code qui créera l’entrée dans le menu déroulant.

```
if($match_pid) {  
    $ret .= $this->ShowMenues($match_pid);  
}
```

Si une correspondance est trouvée, on rappelle la fonction récursivement pour générer les sous-menus de la catégorie.

```
$match_pid=0;  
if ($cat_pid==0) $ret.= "</SELECT>\n<br>";  
return $ret;  
}
```

Enfin, on retourne la chaîne générée, qui n’a plus qu’à être envoyée au navigateur du client.

**FormatResult()** Cette fonction affiche une partie du résultat avec le champ de recherche en gras. Le texte du résultat est tronqué pour être lisible et montrer si possible la première occurrence du terme recherché.

**ShowResults()** Cette fonction effectue la recherche à proprement parler, en utilisant **FormatResult()** pour afficher les résultats. Dans l’extrait de code ci dessous, on peut voir comment la requête SQL est générée.

```
$query="SELECT page_title, page_path,  
SUBSTRING(page_text, 100,6000) AS page_text FROM search_page p ";
```

Ici, on voit qu’on limite déjà la quantité de données retournée par MySQL

```
$this->cat_ids=array_diff($this->cat_ids, array("All"));
if ($no_categories=count($this->cat_ids)){
    $query .= ", search_page_cat pc ";
    $cat_group="AND p.page_id = pc.page_id
    AND pc.cat_id IN
    (". join (',', array_map(intval, $this->cat_ids)) .")
    GROUP BY p.page_id HAVING count(*)=$no_categories ";
}
```

Ici, on met en place la jointure sur les catégories. Elle est stockée dans la variable `$cat_group` qui sera ajoutée à la chaîne de recherche. On utilise la clause “HAVING” pour que les résultats des catégories parentes ne soient pas considérées comme justes dans les catégories filles. Par exemple, si on recherche dans le premier forum : la `cat_id` de la catégorie “Forums” est 4, et la `cat_id` de la catégorie “Premier Forum” est 6. On cherche dans la table `search_page_cat` les entrées qui ont une `cat_id` `x` du type “`x IN (1,6)`”. Dans cet exemple, on voit bien que les résultats du type `x=1` et `x≠6` seront acceptées, ce qui n’est pas bon. Néanmoins, on ne peut pas imposer “`x=1 AND x=6`”, car `x` ne vaut jamais 1 et 6 en même temps. La solution que nous avons adoptée est de récupérer tous les résultats tels que `x=1` ou `x=6` et qui possèdent exactement 2 correspondances. D’une manière plus générale, la clause “HAVING” impose que le nombre de correspondances soit égal au nombre de correspondances imposées. De cette façon, on est sûr que le résultat aura répondu à toutes les demandes.

```
$query .= " WHERE MATCH (p.page_title, p.page_text) AGAINST
    ('".addslashes($this->search_text)."' ) ";
if (!empty($this->search_locale)) {
    $query .= "AND p.page_locale=' "
        .addslashes($this->search_locale)
        ."' ";
}
$query .= $cat_group. " LIMIT 50";
```

Ici, on voit comment la recherche en “FULLTEXT” est effectuée : les mots-clés “MATCH” et “AGAINST” sont utilisés. On voit aussi qu’on choisit de ne récupérer que les 50 premiers résultats. Dans une prochaine version, la page de recherche proposera en bas un lien du type “résultats suivants” qui donnera le reste des pages trouvées. Enfin, l’ordre dans lequel MySQL retourne les résultats dans le cas d’une recherche “FULLTEXT” est par ordre de correspondance décroissante : les pages où le mot-clé est trouvé le plus souvent sont retournées en premier.

**GenPage()** Cette fonction utilise `ParsePath()`, `DisplayForm()` et `ShowResults()` pour générer la page de résultats, et s’occupe de la mise en forme de la page.

**Search()** C’est le constructeur. Il prend `$page_id` comme argument, et teste ensuite l’adresse de la page à laquelle se trouve l’utilisateur (variable `$PHP_SELF`). Si cette page correspond à `/search.php`, alors on affiche les résultats. Sinon, on affiche un champ de recherche simple.

#### 3.2.3 Exemple d’une recherche

Cette classe est instanciée au début de chaque page, sauf la page de recherche elle-même. Elle génère alors un petit champ de recherche et la variable `$page_id` (cachée à l’utilisateur) qui prend la valeur de l’URL de la page courante. Ce champ est ensuite inclus dans les en-têtes de chaque page pour générer un champ de recherche sur toutes les pages du site. Le formulaire est ensuite envoyé vers la page de recherche qui génère les résultats. Pour cela elle crée un objet `search` en lui passant en paramètre la variable `$page_id`, qui est donc l’adresse de la page d’où provient la demande de recherche.

Les données sont ensuite envoyées à la page `search.php` qui instancie la classe `search` avec en paramètre l’url de la page précédente. Cette url est ensuite analysée par la méthode `parse_path()` qui tente d’en extraire les catégories demandées ainsi que la langue demandée par l’utilisateur. De cette façon, un utilisateur qui utiliserait la fonction de recherche à partir d’un certain manuel de la section documentation ne chercherait par défaut que dans ce manuel précis. De même, un utilisateur se servant de cette fonction de recherche à partir des forums ne chercherait par défaut que dans les forums.



La page de recherche lui est alors affichée, et elle est divisée en deux parties. La partie supérieure contient une série de menus déroulant présentant les catégories et sous-catégories qu'il a choisi, avec juste au dessous le texte recherché, et un bouton pour relancer la recherche. Si l'utilisateur change de catégorie, la page se recharge pour lui présenter la nouvelle catégories et un menu déroulant contenant les sous-catégories de son nouveau choix.

La partie inférieure affiche le nombre de résultats trouvés, et affiche ensuite une ligne en gras avec le titre de la page trouvée, suivi d'un extrait de la page commençant à la première occurrence du mot recherché. Dans cet extrait, tous les mots recherchés apparaissent en gras. En cliquant sur le titre de la page trouvée, l'utilisateur est envoyé vers la page correspondante.

Dans la figure 3.2 page suivante, vous pouvez voir une copie d'écran montrant la page de recherche et une recherche en cours.

#### 3.2.4 Les scripts d'approvisionnement des tables de recherche

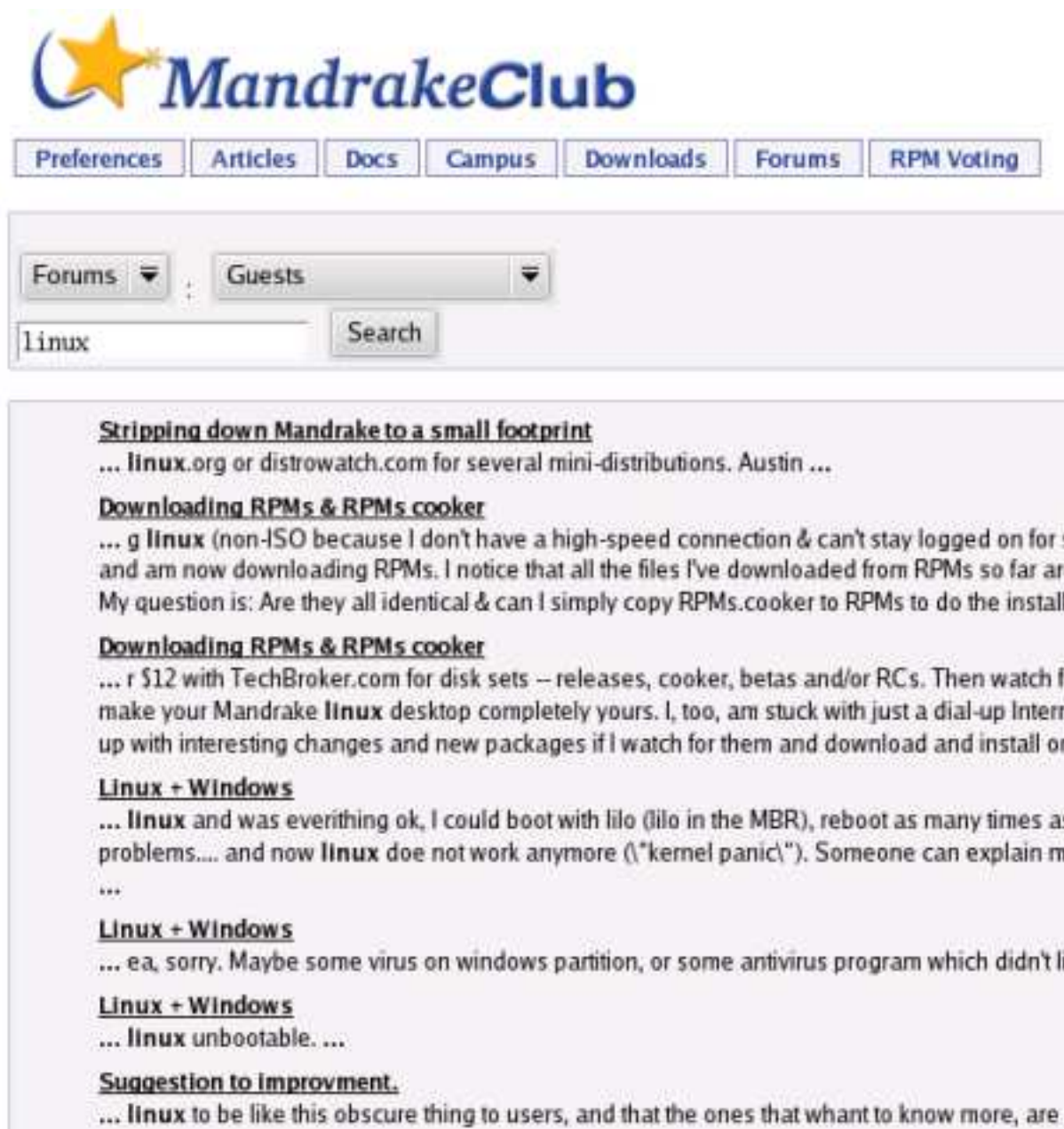
On utilise quatre scripts pour garder à jour la table de recherche :

- Une bibliothèque de fonctions,
- Une fonction pour mettre à jour les articles et les commentaires,
- Une fonction pour mettre à jour les forums,
- Une fonction pour mettre à jour la documentation.

La mise à jour des articles, des commentaires et des forums peut être automatique et être lancée à intervalle réguliers, mais la mise à jour de la documentation nécessite de lancer le script à la main, avec comme paramètre le premier fichier HTML de la documentation en question. Ces scripts utilisent l'interpréteur PHP en ligne de commande.

Si le site ou les catégories n'existent pas, le script les crée avant de migrer les nouveaux articles, commentaires et forums vers la table de recherche. A chaque insertion dans la table `search_page`, les scripts insèrent dans `search_page_cat` les entrées nécessaires pour faire la correspondance. Ces scripts nettoient aussi la base de données pour purger les informations trop vieilles, ou inintéressantes.

Ils peuvent donc tourner à intervalle réguliers et garder la fonction de recherche à jour avec un minimum d'intervention humaine.



The screenshot shows the MandrakeClub website. At the top is the MandrakeClub logo, which consists of a yellow star with a blue outline and the text "MandrakeClub" in blue. Below the logo is a navigation bar with buttons for "Preferences", "Articles", "Docs", "Campus", "Downloads", "Forums", and "RPM Voting". Below the navigation bar is a search interface with a "Forums" dropdown menu, a "Guests" dropdown menu, and a search box containing the text "linux". To the right of the search box is a "Search" button. Below the search interface is a list of forum results. Each result has a title in bold and underlined, followed by a snippet of text. The results are:

- Stripping down Mandrake to a small footprint**  
... linux.org or distrowatch.com for several mini-distributions. Austin ...
- Downloading RPMs & RPMs cooker**  
... g linux (non-ISO because I don't have a high-speed connection & can't stay logged on for ...  
and am now downloading RPMs. I notice that all the files I've downloaded from RPMs so far are ...  
My question is: Are they all identical & can I simply copy RPMs.cooker to RPMs to do the install ...
- Downloading RPMs & RPMs cooker**  
... r \$12 with TechBroker.com for disk sets -- releases, cooker, betas and/or RCs. Then watch for ...  
make your Mandrake linux desktop completely yours. I, too, am stuck with just a dial-up Internet ...  
up with interesting changes and new packages if I watch for them and download and install or ...
- Linux + Windows**  
... linux and was everithing ok, I could boot with lilo (lilo in the MBR), reboot as many times as ...  
problems.... and now linux doe not work anymore (!"kernel panic"). Someone can explain m ...  
...
- Linux + Windows**  
... ea, sorry. Maybe some virus on windows partition, or some antivirus program which didn't li ...
- Linux + Windows**  
... linux unbootable. ...
- Suggestion to improvment.**  
... linux to be like this obscure thing to users, and that the ones that whant to know more, are ...

FIG. 3.2 – Exemple de recherche

### 3.3 La gestion des commentaires

L'objectif était d'ajouter la gestion des commentaires aux pages de manuel proposées sur le site, pour que les membres du club puissent donner leur avis et leurs conseils sur tout ce que propose la documentation : aussi bien la procédure d'installation que l'utilisation au quotidien de la distribution.

Dans phpNuke, les commentaires correspondant aux articles sont stockés dans une table qui s'appelle `nuke_comments`, et qui contient une référence à l'article de la table `nuke_articles` correspondant. De la même façon, il y a une table `poll_comments` qui est utilisée pour les commentaires des sondages, et qui est quasiment la même. Ainsi, le code qui contrôle les commentaires des articles et ceux des sondages sont quasiment les mêmes.

#### 3.3.1 Description générale

L'idée de mon travail était de refondre la gestion des commentaires pour la rendre modulaire, et qu'elle puisse ensuite s'appliquer aussi bien aux articles, aux sondages, et à la documentation.

Les commentaires seront reliés à la page qu'ils commentent par la seule chose qui soit unique : son URL. Ensuite, il faut réécrire quasiment complètement les scripts de gestion des commentaires pour prendre en compte le nouveau système. Pour cela, on utilise un petit groupe de fonctions qui seront dépendantes du type de pages sur lesquelles s'appliquent les commentaires. Cela crée une sorte d'API sur laquelle peut se reposer le script principal. Si on veut appliquer ces commentaires à un nouveau type de page, il suffit alors de modifier ces 2 fonctions pour prendre en compte le nouveau type, et c'est tout. Les deux fonctions qui nécessitent un traitement particulier sont `getPageInfo()` et `showOriginal()`.

**`getPageInfo()`** retourne un tableau associatif contenant l'identifiant de la page (son URL), la langue de la page, son URL privée de la partie correspondante à la langue, et la version de la distribution concernée si elle existe. L'URL sans la partie donnant la langue est utile à cause du caractère multilingue du site. Un article principal peut être traduit en plusieurs langues, mais les commentaires sont associés à l'article principal quelque soit leur langue, de façon à pouvoir répondre dans une langue dans laquelle l'article n'a pas encore été traduit. De même, un membre lisant un article en français peut vouloir lire les

commentaires en anglais aussi si il comprend cette langue.

L'extraction d'informations sur la page se fait en analysant l'URL, et est donc dépendante du type de page affichée. On utilise les fonction d'expression rationnelle de PHP pour trouver les champs qui correspondent, comme montré dans l'exemple de code ci-dessous :

```
if (eregi("docs/", $page_id)) { // It's a manual
    // We may have got an org_page_id instead of page_id
    if (!eregi("docs/(.+)/(.{2})/(.*)", $page_id, $path_values)) {
        if (!eregi("docs/(.+)/(.*)", $page_id, $path_values)) {
            // No, really, we haven't got anything interesting.
            echo "<h1>ERROR parsing path in comments !!</h1>";
            die();
        }
    }
    [...]
} else if (eregi("article.php?", $page_id)) { //It's a story
    [...]
}
```

**showOriginal()** affiche la page courante. C'est une fonction qui est utilisée lorsque l'utilisateur poste un nouveau commentaire. Il a alors l'article sous les yeux et peut copier/coller les lignes auxquelles il veut répondre. Pour afficher la page, il faut donc savoir de quel type de page il s'agit : la documentation est stockée sous forme de fichier texte, et les articles sont stockés dans la base de données. Ici aussi, on utilise les fonction de correspondance à base d'expression rationnelle :

```
if (eregi("^/docs", $page_id)) { // It's a manual
    include("/web/club-downloads/$page_id");
} else { // get the sid and output the article
    [...]
}
```

Ces deux fonctions sont les seules qui dépendent du type de page à afficher, et sont très simples à modifier si on veut appliquer des commentaires à une autre partie du site.

La gestion des commentaires est donc modularisée au maximum, et ne souffre plus de la même duplication de code qu'elle connaissait auparavant.

#### 3.3.2 Description dans le détail

C'est le fichier `comments2.php` qui contient toutes les fonctions nécessaires à la gestion des commentaires : leur affichage, l'ajout d'un commentaire, la réponse à un commentaire précédent, et l'affichage d'un commentaire particulier. La suppression des commentaires est gérée dans une bibliothèque de fonctions réservée aux administrateurs.

##### Affichage de tous les commentaires

Les commentaires sont appelés par l'inclusion dans le script PHP de la page d'un fichier `comments2.php`, puis par l'appel de la fonction `DefaultAction()`, qui lancera l'affichage des commentaires.

Le script lancera alors l'affichage d'une barre de navigation par la fonction `navbar()`, puis appellera la fonction `DisplayComments()` qui affichera récursivement tous les commentaires de la page selon trois modes différents :

- flat : les commentaires sont placés les uns au dessous des autres
- nested : les sous-commentaires sont décalés vers la droite pour marquer le fil de discussion
- thread : les commentaires de premier niveau sont affichés, et le sujet des sous-commentaires est affiché en étant décalé pour refléter le fil de discussion.

La fonction `DisplayComments()` s'appelle donc elle-même et s'appuie sur une autre fonction `ShowComments()` pour faire la génération de code HTML proprement dite, et cela afin de soulager le code et donc d'en améliorer la lisibilité.

Concrètement, c'est l'appel à la fonction `DefaultAction()` qui déclenche le processus. Cette fonction commence par extraire les informations nécessaires de la page grâce à la fonction `getPageInfo()`, puis lance `DisplayComments()`. Cette fonction extrait des informations sur l'utilisateur la liste des langues qu'il comprend, et cherche dans la base les commentaires correspondant à cette liste. Pour chaque commentaire, la fonction appelle `showComment()` qui va afficher le commentaire souhaité et joue sur la valeur de la variable

\$level pour placer le commentaire en décalage ou non, comme le montre l'exemple de code ci-dessous :

```
[...]
while (list( [...] )=sql_fetch_row($commentResult,$dbi)) {
    showComment( [...] );
    if ($level >= 40) {
        // No more than 40 levels deep...
        die("Maximum comments reached");
    }
    switch ($mode) {

        case "nested" :
            $level++;
            DisplayComments($page_id, $tid, $tid, $level);
            $level--;
            break;

        case "flat" :
            DisplayComments($page_id, $tid, $tid, 0);
            break;

        case "thread" :
            $level++;
            DisplayComments($page_id, $tid, $tid, $level, 1);
            $level--;
            break;
    }
    echo "</ul>";
    if($hr) echo "<hr noshade size=\"1\">";
}
[...]
```

On voit bien ici l'appel récursif et l'utilisation de la variable “\$level” pour placer les commentaires. On remarquera aussi que dans le mode “flat”, la variable “\$level” n’est pas

utilisée : tous les commentaires sont placés les uns au-dessous des autres. Les symboles [...] signalent que j’ai supprimé du code pour que l’exemple reste lisible.

Ensuite, la fonction `showComment()` s’occupe de la génération du code HTML, avec les liens pour répondre.

#### Ajout et réponse à un commentaire

L’ajout d’un commentaire se fait par l’utilisation d’un bouton “Répondre” de la barre de navigation. La réponse à un commentaire, elle, se fait en cliquant sur le lien “Répondre” en dessous du commentaire. A ce moment là, le formulaire nous redirige vers l’adresse suivante : `/comments2.php?op=reply`, ce qui indique au script qu’il s’agit d’une réponse. Le script utilise alors la fonction `showOriginal()` pour afficher la page à laquelle répond l’utilisateur, et lui propose ensuite de taper le sujet de son commentaire, le texte du commentaire, et d’en choisir la langue. Par défaut, la langue choisie est celle de l’article. L’utilisateur peut alors avoir une pré-visualisation de son commentaire, ou bien le poster directement. Dans ce cas, le commentaire est inséré dans la table `club_comments`.

Le système permet aussi de visualiser un commentaire unique, et de modérer les commentaires. Modérer un commentaire correspond à lui attribuer une note et une raison d’attribution de cette note (par exemple : “intéressant”, “hors-sujet”, “troll”<sup>1</sup>...).

#### 3.3.3 Conclusion

La nouvelle fonction de gestion des commentaires peut être associée à n’importe quelle page du site. Grâce à elle, il suffit de quelques lignes de code pour créer un système de gestion des commentaires sur n’importe quelle section existante ou à venir. Pour un site de cette taille, et avec un objectif de croissance aussi important, il est primordial de standardiser au maximum les fonctions, et de créer des briques logicielles qu’il suffira alors d’empiler pour construire un site plus grand, offrant plus de possibilités, attirant plus de monde, et donc finalement générateur de plus d’argent pour MandrakeSoft.

---

<sup>1</sup>Dans le jargon des forums de discussion, un “troll” désigne une remarque dont le seul but est de provoquer une discussion passionnée sans fin. Voir : <http://www.linux-france.org/prj/jargonf/T/troll.html> pour plus d’informations.

## 3.4 Autres travaux

### 3.4.1 Portage de [mandrakesoft.com](http://www.mandrakesoft.com) de PHP3 à PHP4

Le site corporatif de la société est <http://www.mandrakesoft.com>. Ce site est actuellement hébergé aux USA. Dans un souci de réduction des coûts, le site va être ramené sur un serveur de MandrakeSoft en France, ce qui permettra d'économiser le coût de l'hébergement de ce site.

Le site a été programmé pour PHP3 et MySQL 3, et le nouveau serveur est équipé de PHP4 et MySQL 4. Il faut donc migrer le site, parce que la compatibilité ascendante d'est pas totale. Pour cela, il a fallu modifier le nom de certains champs de la base de données pour éviter de nouveaux mots réservés par MySQL, et répercuter ces modifications sur le site. PHP4 introduit lui aussi quelques incompatibilités légères dans la gestion des variables et dans certaines fonctions.

La migration du site vers les nouvelles versions de MySQL et de PHP a pris environ deux semaines.

### 3.4.2 Installation du moteur de recherche *RpmFind*

La page <http://rpms.mandrakeclub.com> contenait avant une liste des RPMs disponibles en téléchargement sur le site, mais il n'y avait aucun moyen d'effectuer une recherche parmi ces logiciels. J'ai installé le moteur de recherche qui est utilisé tous les jours sur <http://rpmfind.net> et qui permet d'effectuer des recherches non seulement parmi les titres des RPMs mais aussi parmi les fichiers qu'ils contiennent.

Les RPMs sont classés par catégories sur la page des résultats, et chaque RPM peut être affiché sur une page indépendante qui donne le nom du RPM, son résumé, sa description, la liste des programmes dont il dépend et la liste des fichiers qu'il contient. Le moteur de recherche est en PHP, la base de données est MySQL, et c'est un programme en C qui alimente la base à partir des RPMs disponibles.



#### 3.4.3 Mise en place d'une interface web à l'annuaire LDAP de MandrakeSoft

L'équipe d'administrateurs de MandrakeSoft a pour objectif de mettre en place un serveur LDAP dans le double but de fournir un service d'annuaire aux utilisateurs et de migrer l'authentification des utilisateurs sur les machines de NIS vers LDAP. Le serveur n'est pas encore en place, mais j'ai proposé de réaliser une étude des différentes interfaces web disponibles pour choisir l'interface qui sera proposée aux utilisateurs. Nous avons finalement opté pour DaveDap, qui permet toutes les opérations classiques des serveurs LDAP, et qui supporte parfaitement l'UTF-8 (encodage des caractères dans la base, et donc support des caractères accentués).



Le Projet de Fin d'Études a pour but de préparer au mieux l'élève à son futur emploi, et je crois que ce but a parfaitement été atteint. Les quatre mois que j'ai passé chez MandrakeSoft m'ont fait découvrir cet univers un peu particulier dans lequel vivent les entreprises qui travaillent autour du logiciel libre. Par exemple, plus d'utilisateurs de la distribution MandrakeLinux n'entraîne pas directement plus d'argent dans les caisses de l'entreprise. D'un autre côté, un travail énorme est réalisé par des volontaires, ce qui fait que l'entreprise a à sa disposition énormément plus de ressources qu'elle n'en possède réellement. Et cet avantage énorme ne tient presque qu'à une chose : la réputation de la société. C'est pourquoi l'éthique de travail est placée très haut dans les priorités de l'entreprise : tant qu'elle respectera les contrats explicites, et surtout les contrats implicites du logiciel libre, la communauté des utilisateurs de Linux l'aidera à réaliser la meilleure distribution possible.

Ce qui m'amène à la situation financière de la société : MandrakeSoft arrivera à la fin de la période d'observation en juillet, c'est à dire à une des périodes les plus difficiles pour la société. La distribution MandrakeLinux est éditée environ tous les six mois. Comme la 9.1 est sortie en mars, on peut prévoir la sortie de la 9.2 pour septembre. En juillet, la société sera donc située à environ un mois de la sortie de la prochaine version, ce qui a pour conséquence que les ventes de boîtes de la version précédente seront assez faibles (les clients préférant attendre la version suivante). De plus, l'été est une période assez difficile pour décrocher des contrats avec des entreprises. En juillet, l'entreprise sera donc présentée à l'administrateur judiciaire sous son plus mauvais jour. S'il décide que l'entreprise n'est pas viable et ordonne la liquidation judiciaire, MandrakeSoft sera très certainement reprise par une autre société : il y en a plusieurs qui se sont déclarées intéressées. Ce qui risque d'arriver alors est un virage vers un modèle plus fermé, et la fin de la distribution MandrakeLinux telle qu'on la connaît aujourd'hui. Mais il ne faut pas oublier que toute la distribution est sous licence GPL, et cette licence offre des avantages très intéressants pour ce genre de cas. Puisque toute la distribution est faite de logiciel libre, la communauté des utilisateurs de MandrakeLinux pourra prendre la distribution dans l'état dans lequel elle est maintenant,

changer son nom, et continuer son développement de manière communautaire. C'est ce qu'on appelle "forker" le projet, et c'est très probablement ce qui arrivera si la société est placée en liquidation judiciaire. Bien sûr, il faut s'attendre alors à ce que la distribution subisse le principal inconvénient du modèle de développement communautaire : sa relative lenteur. Chez MandrakeSoft, il y a une trentaine de personnes qui développent à temps plein la distribution, alors que dans un modèle de développement de type Debian<sup>1</sup> le travail est fait pendant le temps libre des participants. Il faudra donc s'attendre à une distribution qui évoluera moins vite, et qui sera peut-être moins à la pointe des technologies libres.

Pour éviter cette situation, et pour sauver la société, MandrakeSoft a mis en place il y a un peu plus d'un an le MandrakeClub. Il a apporté énormément d'argent et de notoriété à l'entreprise, et on peut même dire que c'est grâce au Club que MandrakeSoft est toujours en activité aujourd'hui. Les dirigeants ont bien compris l'importance du Club, et sont en train de recentrer la société autour de lui. La dernière distribution offre dans sa version boîte un abonnement d'un mois au club, et au premier démarrage du système on se voit proposer d'entrer ses identifiants du Club pour que le gestionnaire de programmes prenne automatiquement en compte les logiciels disponibles sur le Club.

Depuis l'été dernier, le site de MandrakeClub s'est vu additionné de nombreuses fonctionnalités, et cela a eu une influence directe sur la quantité d'argent qu'il génère. Il est donc primordial de constamment améliorer le site pour attirer de nouveaux membres, et pour inciter les membres existants à se réabonner lorsque leur abonnement prend fin. C'est dans cet objectif que j'ai effectué mon Projet de Fin d'Études, et nous avons déjà eu des échos dans les forums de membres qui apprécient beaucoup la nouvelle fonction de recherche et les commentaires disponibles sur la documentation.

---

<sup>1</sup>La plus célèbre distribution communautaire

Ce que je retirerai de ce stage est tout d'abord une expérience importante dans le domaine du Logiciel Libre. J'ai pu appréhender en situation réelle les problématiques liées au monde de Linux, qui est très particulier. Le fait de travailler dans un environnement exclusivement composé de logiciels libres m'a permis aussi de découvrir ou d'améliorer ma connaissance des outils de travail et de développement libres. J'ai aussi appris comment se déroule un projet open-source, et comment peut vivre une entreprise dont l'activité est essentiellement tournée vers le logiciel libre.

De plus, d'un point de vue plus technique, j'ai beaucoup amélioré ma connaissance du langage PHP et du développement web. J'ai pu travailler sur un projet de grande envergure, et maintenant je suis tout à fait prêt pour m'attaquer à la réalisation d'un site web en PHP, quelque soit sa taille. J'ai aussi beaucoup appris sur MySQL et sur le langage SQL en général. Je pense que les connaissances que j'ai acquises me permettront de travailler sur un projet de développement web en m'auto-formant si nécessaire, mais sans rester bloqué.

Mon projet professionnel n'est pas de m'orienter vers le développement, mais plutôt vers l'administration système et réseaux. C'est une discipline que j'ai toujours préféré au développement, et c'est encore le cas aujourd'hui. Néanmoins, je suis content d'avoir fait de gros progrès en développement, car ça ajoute au côté généraliste de ma formation. Je sais que même en me tournant vers l'administration système, je serais amené à faire du développement à un moment où à un autre de ma carrière, ou au moins à devoir adapter le programme de quelqu'un d'autre à mes besoins. En développant mes compétences dans ce domaine, je pense avoir effacé un point plus sombre de mes capacités techniques. Grâce à ce stage, j'ai le sentiment d'être plus polyvalent, et cela m'aidera à faire valoir mes compétences d'ingénieur généraliste.