



UNIVERSITE SULTAN MOULAY SLIMANE
FACULTE DES SCIENCES ET TECHNIQUES



Centre d'Étude Doctorales : Sciences et Techniques
Formation doctorale : Mathématiques et Physique Appliquées

THESE

Pour l'obtention le grade de

DOCTEUR

Spécialité : Informatique

**La Mise en Œuvre des Systèmes Intégrés dans le
Système Universitaire : Théoriques et Applications**

Par **Jaouad OUKRICH**

Présentée et soutenue publiquement le **16 Mars 2019** à la **FST de Beni Mellal** devant le jury composé de :

Président	Pr. Cherki DAOUI	FST – USMS – Beni Mellal
Rapporteurs	Pr. Brahim MINAOUI	FST – USMS – Beni Mellal
	Pr. Rachid EL AYACHI	FST – USMS – Beni Mellal
	Pr. Jilali ANTARI	FP – Ibn Zohr – Taroudant
Examineur	Pr. Mohamed FAKIR	FST – USMS – Beni Mellal
Co-Encadrant	Pr. Belaid BOUIKHALENE	FP – USMS – Beni Mellal
Directeur de thèse	Pr. Noureddine ASKOUR	FST – USMS – Beni Mellal

Laboratoire de recherche d'Informatique, FST de Beni Mellal

Dédicace

À mes parents.....

Remerciements

J'exprime toute ma gratitude à mon encadrant Monsieur **Noureddine ASKOUR**, Professeur à l'Université Sultan Moulay Slimane de Beni Mellal (Faculté des Sciences et Techniques) d'avoir accepté de m'encadrer administrativement, et Monsieur **Belaid BOUIKHALENE**, Professeur à l'Université Sultan Moulay Slimane de Beni Mellal (Faculté Polydisciplinaire), pour son aide, son encadrement et ses très précieux conseils.

Je remercie Monsieur **Cherki DAOUI** Professeur à l'Université Sultan Moulay Slimane de Beni Mellal (Faculté des Sciences et Techniques), d'avoir accepté de présider le jury.

Merci également aux personnes qui m'ont fait l'honneur d'accepter de participer à mon jury en tant que des rapporteurs de cette thèse Monsieur **Brahim MINAOUI** Professeur à l'Université Sultan Moulay Slimane de Beni Mellal (Faculté des Sciences et Techniques), Monsieur **Rachid EL AYACHI** Professeur à l'Université Sultan Moulay Slimane de Beni Mellal (Faculté des Sciences et Techniques) et Monsieur **Jilali ANTARI** Professeur à l'Université Ibn Zohr Taroudant (Faculté Polydisciplinaire).

Je remercie également Monsieur **Mohamed FAKIR** Professeur à l'Université Sultan Moulay Slimane de Beni Mellal (Faculté des Sciences et Techniques) d'avoir accepté d'examiner ce travail.

Je tiens à remercier également tous ceux qui m'ont soutenu et qui ont contribué de près ou de loin à l'élaboration de ce travail.

Je remercie toute ma famille et en particulier mes parents et mes sœurs.

Résumé

Les dernières années connaissent une évolution quasi-exponentielle de la quantité d'information, de capacité de stockage et de production due au développement des nouvelles technologies de l'information et de la communication et à l'apparition des nouveaux moyens de partage d'opinions autour d'un service ou d'un produit (via les réseaux sociaux).

Aujourd'hui, les universités sont obligées de changer la manière dont ils gèrent ces évolutions. Elles ont pris conscience de l'importance à la gestion automatique de leurs patrimoines intellectuels, de garder les traces de toutes les transactions effectuées au sein de l'université, de savoir les attentes de différents acteurs y interviennent ainsi de prendre des meilleures décisions au bon moment. Le but à travers ce mémoire de thèse est de proposer à l'université des outils informatiques, reposant sur nos approches. Ces outils doivent offrir un ensemble de fonctionnalités permettant de :

- Répondre à la maxime courante : "la bonne information à la bonne personne au bon moment " ;
- Prendre des meilleures décisions au bon moment par les dirigeants ;
- Favoriser des apprentissages et des améliorations des services aux usagers ;
- Améliorer la visibilité des auteurs et des ouvrages ;
- Analyser les attentes des différents acteurs de l'université ;
- Analyser rapidement et efficacement des données assez volumineuses provenant de plusieurs et divers sources d'informations.

Notre travail se décompose de deux axes principaux (5 chapitres). Dans le premier axe (en deux chapitres), nous présenterons le cadre institutionnel de l'étude et nous élaborerons une méthodologie d'analyse de l'existant et des besoins et la planification du projet. Tandis que dans le deuxième axe (chapitre 3,4 et 5), nous proposerons une démarche détaillée de l'étude de la sélection des outillages open source à savoir le Système Intégré de Gestion de Bibliothèque (SIGB), le Progiciel de Gestion Intégré (PGI) et le système d'analyse de données massives les plus adaptés possible à l'université ainsi que nous donnerons une description des méthodes d'analyse d'opinion. Dans ce dernier axe, nous présenterons une description détaillée des approches proposées et l'implémentation des outils choisis. Le fruit de ce travail consiste dans la valorisation des rôles des SIGB, des ERP et des systèmes de Big Data dans la meilleure gestion des documents d'un établissement universitaire public.

Mots clés : Analyse d'opinion, données massives, Planification, Progiciel de Gestion Intégré, Système Intégré de Gestion de Bibliothèque, université.

Abstract

Recent years have seen an almost exponential evolution in the quantity of information, storage capacity and production because of the revolution of new information and communication technologies and the emergence of new sharing opinions tools about a service or product. Today, universities have had to change the way they manage these developments. They have get the importance of automatic management of their intellectual Knowledge, of keeping track of all the transactions carried out within the university, of knowing the expectations of different actors, and thus of making better decisions at the right time .

The purpose of this work is to propose some tools to the university, based on our approaches. These tools must offer a set of features presented as follow:

- Answer the current maxim: "the right information to the right person at the right time"
- Make better decisions at the right time by the leaders;
- Promote learning and improve user services;
- Improve the visibility of authors and books;
- Quickly and efficiently, analyze fairly large data.

Our work is divided into two main axis (5 chapters). Firstly (chapters 1 and 2), we will present study's institutional framework, and will develop a methodology for analyzing the existing and the needs. Secondly (chapters 3 to 5), we will propose a detailed approach to the study of the open source tools Benchmarking namely the Integrated Library Management System (ILMS), an Enterprise Resource Planning (ERP) and a Big Data analysis system the most suitable possible to the university as well as we will give a description of the opinion analysis methods. In this last axis, we will give a detailed description of the proposed approaches and finally the implementation of the chosen tools. The result of this work is the role enhancement of ILMS, ERP and Big Data systems to improving documents management of a public Moroccan university institution.

Key Word: Big Data, Enterprise Resource planning, ILMS, Integrated system, opinion analysis, University.

ملخص

شهدت السنوات القليلة الماضية تزايدا ملحوظا ومتسارعا في كمية المعلومات المنتجة وبالتالي الزيادة في قدرة تخزينها ومعالجتها، وذلك نتيجة لتطور تكنولوجيا المعلومات والاتصالات وظهور وسائل جديدة لتبادل الآراء حول الخدمات أو المنتجات عبر مواقع التواصل الاجتماعي (الفيسبوك، التويتر...) مما أدى بالمؤسسات الجامعية الى التفكير في تغيير الطرق التي تدبر بها أنشطتها، فقد أدركت مدى أهمية تدبيرها لموروثها الثقافي، وكذا نيتها اعتماد أنظمة معلوماتية يمكنها من تتبع جميع الإجراءات والعمليات المنجزة أو المبرمجة ERP، معرفة انتظارات مختلف الفاعلين من طلبة، أساتذة، موظفين، أعوان وكذلك مختلف الشركاء وذلك بغية تحسين مردودية المؤسسة و اتخاذ قرارات حاسمة في وقت مناسب.

الهدف من هذه الأطروحة هو اقتراح أنظمة معلوماتية متكاملة في النظام الجامعي، والتي تسمح بما يلي:

- تمكين أصحاب القرار بالمؤسسات الجامعية من اتخاذ قرارات جيدة في وقت مناسب.
- توفير معلومات جيدة للطلبة في الوقت المناسب.
- تعزيز التعلم وتحسين الخدمة للمستعملين.
- تحليل انتظارات مختلف الفاعلين داخل الجامعة.
- تحليل قاعدة بيانات ضخمة في وقت جيد وبجودة عالية.

وتنقسم مساهمتنا الى محورين (5 فصول): في المحور الأول (فصلين)، سنقدم الإطار المؤسسي لهذا البحث ثم سنقوم بتحليل احتياجات الجامعة من النظم الالكترونية عبر جرد لمختلف الأنظمة المتواجدة حاليا عن طريق الملاحظة والبحث الميداني. بينما في المحور الثاني (3 فصول)، سننجز دراسة تفصيلية لاختيار الأدوات مفتوحة المصدر (Open source) خصوصا فيما يتعلق بتسيير المكتبات الجامعية (ILMS)، تسيير موارد الشركات (ERP)، تقنيات معالجة البيانات الكبيرة (Big Data)، وكذلك سنقدم وصفاً لطرق تحليل الرأي. ثم سنقدم وصفا مفصلا للنهج المقترحة وسنتناول الجانب العملي للأدوات التي تم اختيارها سالفا وتقديم بعض النتائج مصاحبة ببعض الاستنتاجات وعرض بعض تصورات لاستكمال هذا البحث في المستقبل. تتجلى نتائج هذا البحث في تعزيز دور الأدوات السالفة الذكر في التسيير الجيد للوثائق ولأرشيف المؤسسات الجامعية المغربية.

كلمات مفتاح: تسيير موارد الشركات، الجامعة، أنظمة معلوماتية متكاملة، تسيير المكتبات الجامعية، البيانات الكبيرة، تحليل الرأي.

Table des matières

Table des matières	vi
Sigles et abréviations utilisées	x
Liste des tableaux.....	xii
Liste des figures	xiii
Liste des publications.....	xvi
Articles publiés :	xvi
Communications dans des congrès internationaux :	xvi
Introduction Générale	1
a) Contexte et motivation	1
b) Problématiques	1
c) Propositions et contributions	2
d) Organisation du mémoire	2
Chapitre 1. Présentation du cadre institutionnel, analyse de l'existant et planification du projet.....	4
1.1 Introduction	4
1.2 Présentation du cadre institutionnel.....	4
1.2.1 Présentation de l'USMS de Beni Mellal.....	4
1.3 Recensement des systèmes informatiques intégrés à l'échelle de l'université...5	
1.3.1 SIGB	5
1.3.2 ERP ou PGI.....	6
1.3.3 APOGEE.....	6
1.3.4 Moodle	6
1.3.5 Applications personnelles et autres sources d'information	6
1.4 Présentation et analyse des résultats des enquêtes.....	6
1.4.1 Techniques d'enquête et mode d'administration.....	6
a) Enquête par entretien	7
b) Enquête par questionnaire	7
1.4.2 Résultats d'enquête par questionnaire	7
a) Analyse descriptive (analyse fréquentielle)	7
b) Analyse descriptive (Analyse de la moyenne)	16
c) Classification automatique non supervisé.....	16
d) Analyse avec ACM.....	17

e) Résultats de l'enquête en utilisant le nuage de mots	19
f) Résultat de l'enquête en utilisant le diagramme de Pareto	19
1.5 Planification et budget prévisionnel du projet.....	22
1.5.1 Phases de réalisation du projet	23
a) Phase de définition et du montage	23
b) Planification en utilisant le réseau PERT et Gantt	25
c) Suivi et évaluation du projet.....	29
1.6 Conclusion.....	29
Chapitre 2. De la base de données relationnelle à l'analyse de données massives (Big Data).....	31
2.1 Introduction	31
2.2 Définition d'un Système d'Information.....	31
2.3 Historique des SI dans les Universités Marocaines	33
2.4 Méthodes de conception d'un système d'information.....	34
2.5 Système de gestion de Base de données relationnelle et langage SQL	34
2.5.1 Base de Données Relationnelle	34
2.5.2 SGBDR	34
2.5.3 SQL.....	35
2.6 Modèle multidimensionnel.....	35
2.7 Architecture et fonctionnement d'un système décisionnel	36
2.7.1 Entrepôt de données et base métiers.....	36
2.7.2 Outils de reporting	37
2.8 Analyse de données, outils statistiques et forage de données (Data Mining) .	38
2.8.1 Statistique descriptive	38
2.8.2 Statistique exploratoire	39
a) Méthodes exploratoires de corrélation.....	39
b) Méthodes exploratoires de segmentation par apprentissage non supervisé	41
c) Méthodes exploratoires de segmentation par apprentissage supervisé.....	43
2.9 Text mining, méthodes d'analyse des sentiments/d'opinions et outils de visualisation des données	44
2.9.1 Text mining	44
2.9.2 Analyse d'opinion ou des sentiments et outils de visualisation de texte...	48
a) Analyse d'opinion ou de sentiments.....	48
b) Outils de visualisation de texte : Nuage de mots.....	50

2.10	Limites du système d'information relationnel.....	53
2.11	NoSQL et Big Data.....	53
2.12	Conclusion.....	56
Chapitre 3. Benchmarking des outillages.....		58
3.1	Introduction	58
3.2	Benchmarking de l'outillage SIGB	58
3.2.1	Solutions gratuites	58
3.2.2	Solutions propriétaires	64
3.2.3	Critères de choix et résultat.....	65
3.3	Benchmarking de l'outillage des ERP	68
3.3.1	Solutions gratuites	68
3.3.2	Solutions propriétaires	71
3.3.3	Critères de choix et résultat.....	71
3.4	Benchmarking de l'outillage d'analyse de données massives (Big Data)	72
3.5	Conclusion.....	74
Chapitre 4. Contributions		76
4.1	Introduction	76
4.2	Nouvelle stratégie de management documentaire de la bibliothèque local de l'USMS	76
4.2.1	Motivation et objectifs	76
4.2.2	Indicateurs de performance de la bibliothèque	78
4.2.3	Méthode proposée	81
4.3	Module d'analyse de référentiel de connaissances (Knowledge Management).....	82
4.3.1	Motivation.....	82
4.3.2	Cas d'utilisation de référentielle connaissance	83
4.4	Module d'analyse des sentiments/opinions.....	85
4.4.1	Motivations	85
4.4.2	Méthodes utilisées pour analyser les sentiments : cas des étudiants	85
	a) En utilisant les nuages de mots.....	85
	b) Analyse d'opinion en se basant sur le lexique des opinions	86
4.5	Approche proposée pour l'intégration d'un Progiciel de gestion intégré ERP dans le management de l'université	86
4.5.1	Motivation et objectifs	86

4.5.2	Méthode de mise en œuvre de l'ERP	87
4.6	Approche proposée pour l'analyse de données massives de l'université	87
4.6.1	Motivation et objectifs	87
4.6.2	Méthode de mise en œuvre d'un Système intégré pour le Big Data	87
4.7	Conclusion.....	88
Chapitre 5.	Implémentation et résultats	90
5.1	Introduction	90
5.2	Réalisation du module SIGB (PMB)	90
5.2.1	Acquisition et installation des logiciels de management de la bibliothèque.....	90
5.2.2	Évaluation de l'activité en bibliothèque de l'USMS de Beni Mellal en utilisant le SIGB PMB et l'outil décisionnel power designer BI.....	91
5.3	Application de la référentielle connaissance	96
5.4	Réalisation de l'ERP de l'USMS.....	99
5.5	Réalisation du module Analyse des sentiments	100
5.6	Conclusion.....	102
Conclusion générale et perspectives		103
Bibliographie		105
Glossaire.....		109
Annexe		110
Annexe 1: Quelques SGBD.....		110
Annexe 2 : TF-IDF.....		110
Annexe 3 : Niveau de satisfaction		111
Annexe 4 : Configuration d'EasyPHP		112
Annexe 5 : Critère de choix de la version de PMB et paramètres.....		113
Annexe 6 : Fiche d'importation des exemplaires		114
Annexe 7 : Enquête de satisfaction des étudiants.....		115

Sigles et abréviations utilisées

Terme	Signification
ACM	Analyse de Correspondances Multiples
ACP	Analyse en Composante Principale
ANN	Artificial Neural Network (Réseau de neurones artificiel)
APOGEE	Application Pour l'Organisation et la Gestion des Enseignements et des Étudiants
BD	Base de Données
CAH	Classification Ascendante Hiérarchique
ERP/PGI	Enterprise Resource Planning ou Progiciel de Gestion intégré
EST	École Supérieur de la Technologie
ETL	Extraction, Transformation and Loading (Extraction, transformation et Chargement)
FLSH	Faculté des Lettres et des Sciences Humaines
FP	Faculté Polydisciplinaire
FST	Faculté des Sciences et Techniques
KNN	K-Nearest Neighbor (K plus proches voisins)
MERISE	Méthode d'étude et de Réalisation Informatique pour les Systèmes d'Entreprise
MOA	Maîtrise d'Ouvrage
MOE	Maîtrise d'Œuvre
OLAP	OnLine Analytical Processing
OMG	Object Management Group
OPAC	Online Public Access Catalogue
PMB	PhpMybiBli

Terme	Signification
RI	Recherche d'information
SGBD	Système de Gestion de Base de Données
SGBDR	Système de Gestion de Base de Données Relationnelle
SI	Système d'Information
SIG	Système d'Information Géographique
SIGB	Système Intégré de Gestion de Bibliothèque
SIRH	Système d'Information en Ressources Humaines
SVM	Séparateurs À Vaste Marge ou Machines à Vecteurs de Support
TALN	Traitement Automatique de la langue naturelle
TIC	Technologie de l'Information et de la Communication
UML	Unified Modeling Language (Langage de modélisation Unifié)
UNIMARC	Universal Machine Readable Cataloging
USMS	Université Sultan Molay Slimane

Liste des tableaux

Tableau 1: Analyse descriptive (moyenne)	16
Tableau 2: Nombre des enquêtés dans chaque classe en utilisant la CAH	16
Tableau 3: Nombre des enquêtés dans chaque classe en utilisant la méthode KMeans	17
Tableau 4 Résultat de l'analyse avec la loi de Pareto	20
Tableau 5 Tableau des antécédents des tâches	26
Tableau 6 Choix d'une technique d'analyse de données descriptive.....	39
Tableau 7 Critères booléens dans la recherche d'information utilisant le moteur de recherche Google	45
Tableau 8 Quelques applications utilisant l'analyse des sentiments	49
Tableau 9 Liste des SIGB gratuites.....	59
Tableau 10 Logiciels open sources de la gestion de bibliothèque	60
Tableau 11 Langues supportées, documentation et taille du logiciel	61
Tableau 12 Mode de développement des SIGB.....	61
Tableau 13 Liste des logiciels de SIGB propriétaires	64
Tableau 14 Schéma de système de notation défini par Tirstan Muller	66
Tableau 15 Principales caractéristiques des cinq logiciels open sources	66
Tableau 16 Évaluation des cinq logiciels open sources de la gestion de bibliothèque	67
Tableau 17 Quelques caractéristiques des ERP open sources	68
Tableau 18 : Outil power BI.....	82
Tableau 19 Nombre de prêt par année et par mois entre 2013 et 2018.....	93
Tableau 20 Notices bibliographiques par ordre décroissant de nombre de fois prêtés	95
Tableau 21 Exemples des documents qui ne sont jamais prêtés.....	96
Tableau 22 Base de données des mémoires.....	97
Tableau 23 : Liste des SGBD existe sur le marché.....	110
Tableau 24: Niveau de satisfaction concernant les heures d'ouverture, la réception du personnel et l'état des locaux de la bibliothèque.....	111

Liste des figures

Figure 1: Répartition des utilisateurs par genre	8
Figure 2 Répartition du statut des répondants.....	8
Figure 3 Répartition des répondants par faculté d'appartenance.....	8
Figure 4 Répartition des répondants par tranche d'âge	8
Figure 5 Répartition de l'état de bourse	9
Figure 6 Domaine d'étude.....	9
Figure 7 Fréquence de la visite à la bibliothèque.....	9
Figure 8 Niveau de satisfaction des services offerts par la bibliothèque traditionnelle	10
Figure 9 Catégories des documents les plus utilisés	10
Figure 10 Obtention des documents demandés.....	10
Figure 11 Principales fonctionnalités souhaitées à informatiser plus rapidement	11
Figure 12 Temps passé à prêter un livre.....	11
Figure 13 Temps passé à la bibliothèque.....	12
Figure 14 Raisons de visiter la bibliothèque.....	12
Figure 15 Aide principale pour rechercher des documents	12
Figure 16 Enquêtés et ouvrages recommandés par le professeur	13
Figure 17 Enquêtés et l'habitude d'acheter des documents.....	13
Figure 18 Lecture des documents d'autres disciplines	13
Figure 19 Possession d'un ordinateur personnel.....	14
Figure 20 Méthode préférée pour accéder à Internet.....	14
Figure 21 Utilisation de la technologie de l'information et de la communication	14
Figure 22 Utilisation de la documentation électronique.....	15
Figure 23 Activités à la bibliothèque.....	15
Figure 24 Dendrogramme de classes des enquêtés	17
Figure 25 Étude de la relation formation, connaissances informatiques dans la recherche du document souhaité.....	18
Figure 26 Résultat de l'enquête avec ACM.....	18
Figure 27 Nuage de mots extrait de l'analyse de l'enquête	19
Figure 28 Diagramme de Pareto (analyse prioritaire)	21
Figure 29 Cycle de vie d'un projet	22
Figure 30 Analyse SWOT.....	24
Figure 31 Diagramme 7M ou d'Ishikawa	24

Figure 32 Caractéristiques du projet.....	25
Figure 33 Diagramme de Gantt du projet	28
Figure 34: Définition de l'information selon Balmisse.....	31
Figure 35 : Décomposition systémique de l'entreprise	32
Figure 36 : Phases de réalisation d'un système d'information	33
Figure 37 Modèle multidimensionnel en étoile	35
Figure 38 Architecture d'un Système d'Information décisionnel	36
Figure 39 Jeux de données d'ACP	40
Figure 40 Algorithme de Kmean.....	41
Figure 41 Processus de textmining et de visualisation des textes en nuage de mots	46
Figure 42 : Processus des transformations subits par le corpus de texte.....	47
Figure 43 Processus d'analyse d'opinion	49
Figure 44 Quelques cas d'utilisation des nuages de mots	50
Figure 45 Différents approches de visualisation du nuage de mots	52
Figure 46 Les 10 Vs de Big Data	54
Figure 47 Processus de Benchmarking des outillages.....	58
Figure 48. Interface utilisateur OPAC pour ABCD	62
Figure 49. Interface utilisateur OPAC pour OpenBiblio	62
Figure 50. Interface utilisateur OPAC pour PMB.....	63
Figure 51. Interface utilisateur OPAC pour Potnia	63
Figure 52. Interface utilisateur OPAC pour Bibliotecaphp	64
Figure 53 Interface Odoo 11	70
Figure 54 Interface d'authentification Dolibarr FP de Beni Mellal	70
Figure 55 Gestion électronique des documents en Dolibarr	71
Figure 56 Architecture Hadoop.....	73
Figure 57 : Management de la bibliothèque	77
Figure 58 Statistiques des exemplaires.....	79
Figure 59 : Statistique de prêt	79
Figure 60 : Statistiques des exemplaires achetés ou créés.....	80
Figure 61 Statistique des lecteurs.....	80
Figure 62 Processus de l'approche proposé.....	81
Figure 63 Cas d'utilisation pour annotation des documents	83
Figure 64 Processus de la génération du nuage de mots des données extraites d'une enquête	85
Figure 65 Processus de mettre en œuvre notre ERP	87

Figure 66 Schéma global du projet.....	88
Figure 67 Fenêtre de connexion au système PMB	92
Figure 68 Nombre de prêts par jour et par tranche horaire de la bibliothèque FP	92
Figure 69 Nombre total des prêts par catégorie (section) depuis 1 septembre 2013 au 1 Juin 2018	93
Figure 70 Nombre des exemplaires par localisation (Consulté le 22 Mars 2019)	94
Figure 71 Nombre d'exemplaire par section (le 22 Mars 2019).....	94
Figure 72 Nuages de mots et diagramme en barplot des mots les plus fréquents [Base de données de la FP]	97
Figure 73 Nuages de mots et diagramme en bar plot des mots les plus fréquents [Base de données des thèses FST département des sciences de la terre].....	98
Figure 74 Nuages de mots et diagramme en bar plot des mots les plus fréquents [Base de données des Mémoires de licence et Master FST département des sciences de la terre]	98
Figure 75 Exemple d'utilisation des mots les plus fréquents	99
Figure 76 Tableau de bord du progiciel ERP proposé pour FP de Beni Mellal	100
Figure 77 Score des opinions analysées d'une enquête de satisfaction	101
Figure 78 Menu déroulant d'EasyPhp	112
Figure 79 Paramétrage du système et de la base de données PMB.....	114
Figure 80 Modèle de création des notices bibliographiques en format .csv	114
Figure 81 Conversion du fichier exemplaires en format Unimarc ISO2709	114

Liste des publications

Articles publiés :

- *A Survey of Users' Satisfaction in the University Library by Using a Pareto Analysis and the Automatic Classification Methods*. IJLIS, vol 6 no.1, pp. 17-36, Web. 2 Apr. 2018. doi:10.4018/IJLIS.2017010102
- *Management Decision System for the Documentary Archives in the Library of a Public Moroccan Institution: Case of Sultan Moulay Slimane University, Beni Mellal*, “International Journal of Computer, Electrical, Automation, Control and Information Engineering”, Vol:9, No:7, 2015. © World Academy of Science, Engineering and Technology; Available online at <http://waset.org/Publications/?path=Publications&q=oukrich&search=Search> .

Communications dans des congrès internationaux :

- *Word clouds for opinions analysis of the University readers', the 4th International Conference on Business Intelligence (CBI'18), Beni Mellal, Morocco, April 25-27 2018.*
- *Opinion analysis of the university library readers', 4th International Conference on Big Data Analysis And Data Mining, September 07-08, 2017 | Paris, France*
- *A comparative study of the open source digital library software, the International Arab Conference on Business Intelligence (ACIT'16), Beni Mellal, Morocco, December 6-8 2016.*
- *Contribution to Measure the User's Satisfaction of the Public Moroccan Libraries, 2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV), vol. 00, no. , pp. 423-430, 2016, doi:10.1109/CGiV.2016.95*
- *Contribution to the automatic categorization of documents , JDOC'15, Beni Mellal, Morocco, Mars 26-28, 2015.*
- *A survey on informatisation of university library in three Maghreb countries : Morocco, Algeria and Tunisia , 1^{ère} Édition des Journées Doctoriales, organisée à l'occasion de la commémoration de 20^{ème} anniversaire de la FST de Beni Mellal sous thème : Recherche Scientifique : Innovation et Développement, Beni Mellal, Maroc, 10-11 juin 2014.*
- *Best prediction algorithm binarization of document images text , First International Conference on Business Intelligence (CBI'14), Beni Mellal, Morocco, April 29-30 2014.*

Introduction Générale

a) Contexte et motivation

Les dernières années connaissent une croissance remarquable des données produites et stockées au sein des établissements universitaires, ce qui a débouché à une explosion des volumes de données disponible. D'où l'apparition de la notion de Big Data (qui se traduit en français données massives ou données volumineuses). Ce qui a obligé les universités à revoir la manière dont elles gèrent leurs patrimoine intellectuelle (gestion de scolarité, de la bibliothèque, de paiement, d'absence, de congé, d'apprentissage en ligne...) et toutes autres sources d'information externes (micro blogs, email, chat...).

Le but de ce travail et la mise en place des systèmes intégrés adapté à l'environnement universitaire en prenant le cas d'étude l'Université Sultan Moulay Slimane (USMS) de Beni Mellal. Ce choix est motivé par :

- Le statut accordé à l'université comme étant un établissement public doté de l'autonomie administratif, financière et pédagogique ;
- L'absence d'un système d'information intégré au niveau de chaque établissement de l'Université étudié ;
- Le souhait exprimé par les responsables au niveau central ou par chaque établissement de disposer des outils informatiques fiables et à moindre coût permettant un accès fiable et facile à l'information en vue de prendre des meilleures décisions au bon moment ;
- L'existence d'un flux important d'informations non exploitables au niveau de l'université.

b) Problématiques

En fin d'année 2012, les universités Marocaines ont pris conscience de l'importance de la gestion automatique de leur patrimoine intellectuelle (la gestion des documents), de garder les traces de toutes les transactions effectuées ainsi de prendre des meilleures décisions au bon moment. Les enjeux sont variés parmi lesquels on peut citer :

- Problématique liée à la recherche d'information documentaire ;
- Problématique liée à la détermination du besoin en ouvrages de la bibliothèque ;
- Problématique liée au référentiel de connaissances ;
- Problématique liée à l'analyse des sentiments des étudiants et du personnel ;

- Problématique liée à la gestion du processus transactionnelle de l'ensemble de l'université ;
- Problématique lié à la gestion d'une masse de données par les systèmes d'information traditionnels.

c) **Propositions et contributions**

Dans ce travail, nous élaborerons une méthodologie d'analyse de l'existant et des besoins. Ensuite, nous proposerons une démarche détaillée de l'étude de la sélection des outillages à savoir le SIGB, un ERP et un système d'analyse de données massives les plus adaptés possible à l'université ainsi que nous donnerons une description des méthodes d'analyse d'opinion, et finalement l'implémentation des outils choisis.

Nos contributions consiste dans l'application des techniques d'analyse de données d'enquête de satisfaction qui n'existe pas dans la littérature à savoir la loi de Pareto, l'adoption du package du logiciel R dans l'analyse des sentiments, des études de benchmarking des outillages des SIGB et des ERP, et finalement notre contribution dans la proposition d'un modèle d'analyse de données massives de l'université utilisant le Hadoop HDFS.

d) **Organisation du mémoire**

La thèse est organisée en 5 chapitres :

Le premier chapitre de ce mémoire de thèse concernera la présentation du cadre institutionnel de l'étude, l'analyse de l'existant et la planification de ce projet informatique.

Le deuxième chapitre traitera l'état de l'art sur les Systèmes d'Information (SI) dont on présentera l'évolution du SI de la base de données relationnelle au Big Data.

Le troisième chapitre sera dédié à l'étude Benchmarking de l'outillage des outils de la mise en œuvre du projet.

Le quatrième chapitre présentera nos contributions : réalisation d'un système décisionnel pour le management de la bibliothèque, pour la capitalisation des connaissances, pour l'analyse des opinions/sentiments et enfin un système dédié à l'analyse de données massives.

Le cinquième chapitre traitera l'implémentation et présentera les résultats des différentes contributions. Ce mémoire se terminera par un ensemble de conclusions et de perspectives des travaux en cours et à venir.

Chapitre 1. Présentation du cadre institutionnel, analyse de l'existant et planification du projet

Chapitre 1. Présentation du cadre institutionnel, analyse de l'existant et planification du projet

1.1 Introduction

Ce premier chapitre de ce mémoire indiquera non seulement les techniques d'analyse de l'existant, mais aussi présentera l'instrument utilisé pour planifier et contrôler le suivi et la faisabilité technique de ce travail de recherche (projet informatique). Ainsi, ce chapitre tournera autour de quatre axes principaux :

- Présentation du cadre institutionnel de l'étude et recensement des différents systèmes intégrés au sein de l'Université ;
- Présentation des méthodes d'enquêtes et d'analyses des données ;
- Vérification du choix du sujet par l'analyse des résultats des enquêtes ;
- Présentation des outils de suivi et d'évaluation du projet.

1.2 Présentation du cadre institutionnel

1.2.1 Présentation de l'USMS de Beni Mellal

L'Université Sultan Moulay Slimane, est un établissement de l'enseignement supérieur public, créé en 2007, est constituée des sept établissements avec des implantations dans les villes de Beni Mellal, Khouribga et Khenifra, l'USMS se compose de :

- La Faculté des Lettres et Sciences Humaines (FLSH) Béni Mellal ;
- La Faculté des Sciences et Techniques (FST) Béni Mellal ;
- La Faculté Polydisciplinaire (FP) Béni Mellal ;
- La Faculté Polydisciplinaire Khouribga ;
- L'École Nationale des Sciences Appliquées (ENSA) Khouribga ;
- L'École Supérieure de Technologie (EST) Khénifra ;
- L'École Supérieure de Technologie(EST) Béni Mellal.

❖ *Quelques chiffres de l'établissement de l'USMS (d'après le site officiel de l'USMS de Beni Mellal, consulté le 22/03/2019) :*

- 07 Établissements ;
- 36 178 Étudiants ;
- 298 Étudiants étrangers ;
- 203 Staff administratif et technique ;
- 549 Enseignants chercheurs ;
- 802 Doctorants ;
- 58 Laboratoires ;
- 07 centres.

Le service d'information et documentation de l'USMS, à travers ses bibliothèques, dispose de quatre (4) types de ressources qui sont :

- Les ressources humaines ;
- Les ressources financières ;
- Les ressources matérielles ;
- Les ressources documentaires.

1.3 Recensement des systèmes informatiques intégrés à l'échelle de l'université

1.3.1 SIGB

L'USMS dispose d'un fond documentaire en partie constituée de documentation, organisée et constituée de livres, thèses, mémoires et rapports de stages : c'est le dépôt institutionnel.

En fin de l'année 2012, la Faculté Polydisciplinaire, la Faculté des Sciences et Techniques de Béni Mellal et l'École Supérieure de la Technologie ne dotent pas d'aucun Système intégré de Gestion de Bibliothèque, par contre la FLSH adopte un système de gestion de sa bibliothèque, mais il est limité (manque de statistiques, savoir la disponibilité des ouvrages, problème de réservation à distance...). C'est pour cette raison, nous avons adopté un questionnaire pour étudier la satisfaction des lecteurs, des bibliothécaires, le corps professoral au travers les services de la bibliothèque traditionnelle ainsi que de savoir leurs attentes et leurs besoins.

1.3.2 ERP ou PGI

À la fin de l'année 2012, aucun des établissements Marocaines relevant à l'Université Sultan Moulay Slimane n'adoptent aucun Progiciel de Gestion Intégré en raison du prix cher du produit ou par effet de résistance des personnels aux changements.

1.3.3 APOGEE

Est un logiciel de gestion de la scolarité. Il apporte des réponses précises en matière de clarification de l'offre de formation, d'amélioration de l'accueil des étudiants, de gestion de la scolarité et de pilotage de l'établissement.

1.3.4 Moodle

Moodle est l'abréviation en anglais de Modular Object-Oriented Dynamic Learning Environment qui se traduit en français Environnement orienté objet d'apprentissage dynamique modulaire. Il est en PHP, libre et distribuée sous la Licence publique générale GNU.

C'est une Plateforme Pédagogique en ligne de l'université Sultan Moulay Slimane, outre la création de cours à l'aide d'outils intégrés (ressources et activités) à l'usage des formateurs, Moodle offre des possibilités d'organisation des cours par filières (catégories et sous-catégories) ,et qui lui donnent également des caractéristiques propres à la mise en place de dispositifs complets d'enseignement.

1.3.5 Applications personnelles et autres sources d'information

Les différents établissements de l'Université adoptent encore des applications usuelles utilisées dans l'informatique de gestion à savoir la suite bureautique de Microsoft office (Excel, Access, Word...), Visual Timetabling ainsi que d'autres applications spécifiques.

1.4 Présentation et analyse des résultats des enquêtes

1.4.1 Techniques d'enquête et mode d'administration

Dans cette étude, nous avons adopté les deux types d'enquêtes utilisées dans les sciences sociales et humaines suivantes :

- Enquête par entretien ou par observation ;
- Enquête par questionnaire.

a) **Enquête par entretien**

Ce mode d'enquête aidera à déterminer les besoins des dirigeants, le corps professoral ainsi que les lecteurs et les bibliothécaires. Nous avons posé la question suivante : Quel est votre avis sur le projet d'informatisation de la bibliothèque ?

Nous avons enquêté 34 personnes (2 bibliothécaires, 2 professeurs et 30 étudiants).

L'analyse des enquêtes par entretien nous a permis de déduire les résultats suivants :

- Certains bibliothécaires expriment leurs besoins de l'informatisation du fond documentaire, tandis qu'il y a des autres qui sont en état de résistance au changement.
- Les étudiants expriment l'importance de l'informatisation des bibliothèques relevant à l'Université.

b) **Enquête par questionnaire**

C'est le mode d'enquête le plus utilisé, et qui permet de récolter l'avis de plusieurs individus à propos un sujet ou une problématique.

Pour une bonne conduite de cette étude d'enquête (voir Annexe 7), nous avons la distribuée aux lecteurs des bibliothèques de l'USMS dans différentes disciplines (physique, informatique, mathématiques, droit privé français, économie, géologie, biologie, géographie, histoire, arabe, éducation islamique, française et anglaise), afin d'attirer davantage des répondants et d'améliorer ainsi la qualité des données. Et pour mesurer le niveau de satisfaction des lecteurs des services présentés par la bibliothèque traditionnelle, nous avons opté l'échelle de Likert [1] à quatre niveaux allant de 1 (Pas satisfait) à quatre (très satisfait). Le questionnaire était anonyme et volontaire. Pour analyser des données recueillies, nous avons utilisé des outils de traitement de données à savoir Microsoft Excel et notamment son outil de statistique XLSTAT.

Dans cette section, nous avons traité les données recueillies en utilisant l'analyse descriptive, le diagramme de Pareto ainsi que les méthodes de classification non supervisées telles que KMeans, CAH et la méthode d'analyse de correspondance multiple (ACM) en vue d'analyser les corrélations entre les différentes opinions des répondants.

1.4.2 Résultats d'enquête par questionnaire

a) **Analyse descriptive (analyse fréquentielle)**

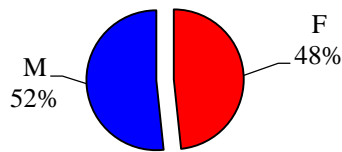


Figure 1: Répartition des utilisateurs par genre

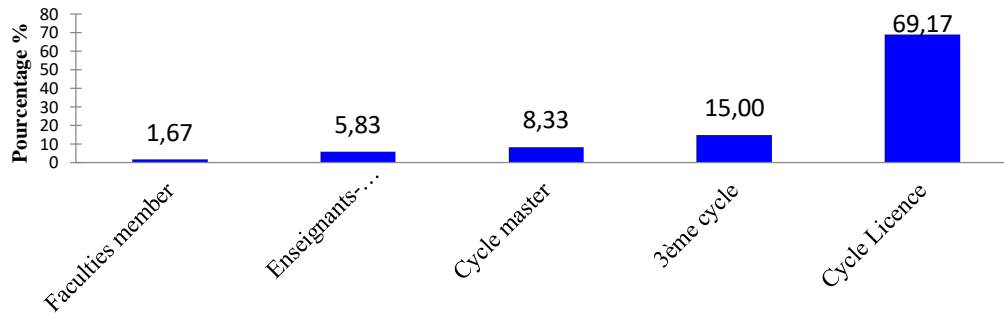


Figure 2 Répartition du statut des répondants

Au total, 980 réponses recueillent 52% d'hommes et 48% de femmes (cf. figure 1) 6% d'enseignants chercheurs ou des contractuels, 69% d'étudiants appartiennent au premier cycle, 8% d'étudiants en master, 15% sont inscrits en 3^{ème} cycle et environ 2% sont des administrateurs (cf. figure 2).

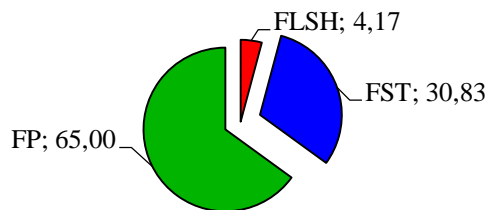


Figure 3 Répartition des répondants par facultie d'appartenance

La population enquêtée est de 65% qui appartiennent à la FP, 31% à la FST, et seulement 4% appartiennent à la FLSH (cf. figure 3)

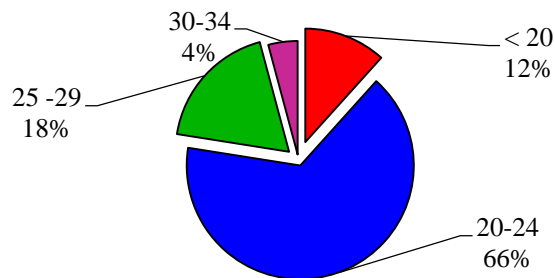


Figure 4 Répartition des répondants par tranche d'âge

La figure 4 exprime que la majorité des répondants appartiennent au tranche d'âge 20-25 ans (66%), 18% ont entre 25-30 ans, 12% ont moins de 20 ans et seulement 4% au groupe d'âge 30-35 ans. Cela implique que cette première catégorie dominante exprime davantage leurs besoins via cette enquête par rapport aux autres tranches d'âge.

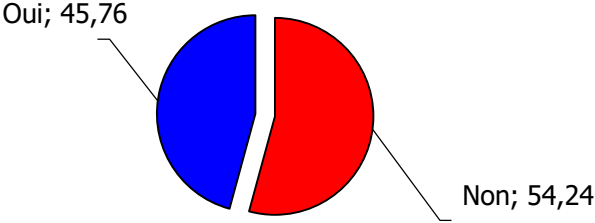


Figure 5 Répartition de l'état de bourse

La figure 5 indique qu'environ la moitié des répondants (54%) ne bénéficient pas de bourse d'étude.

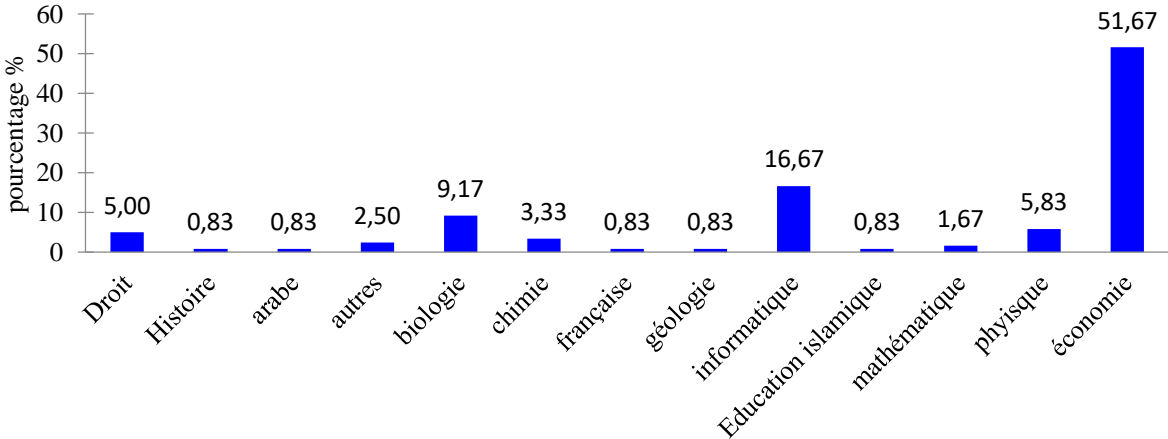


Figure 6 Domaine d'étude

La figure 6 ci-dessus montre que les quatre principaux domaines d'études des enquêtés participant à cette enquête sont l'économie (52%), l'informatique (17%), la biologie à 9%, et la science physique à 6%.

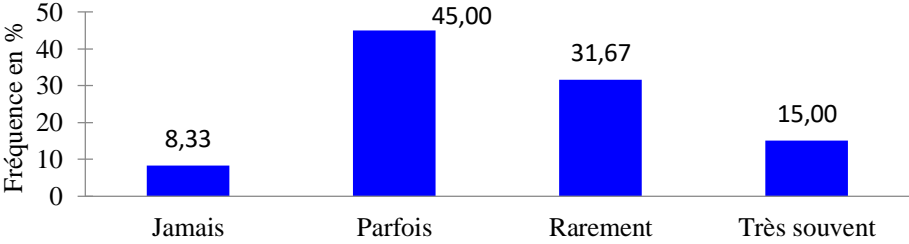


Figure 7 Fréquence de la visite à la bibliothèque

La figure 7 montre que 45% des enquêtés visitent parfois les bibliothèques, 32% l'utilisent rarement, 15% l'utilisent très souvent, et seulement 8% d'entre eux n'utilisent pas la bibliothèque.

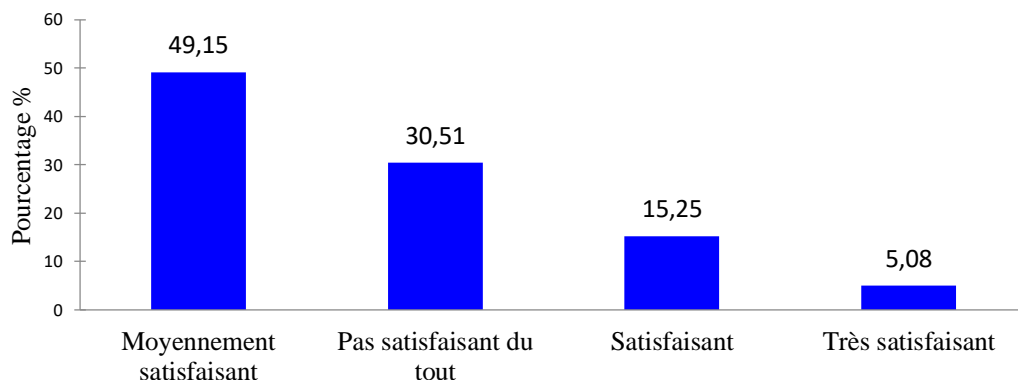


Figure 8 Niveau de satisfaction des services offerts par la bibliothèque traditionnelle

La figure 8 montre que 49% des enquêtés sont moyennement satisfaits des services rendues par la bibliothèque traditionnelle ; 31% ne sont pas satisfaits, 15% sont en état de satisfaction, tandis que seulement 5% sont très satisfaits.

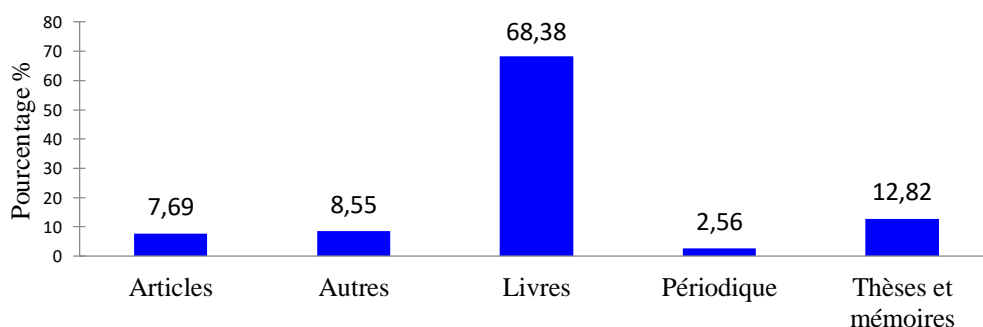


Figure 9 Catégories des documents les plus utilisés

La figure 9 démontre que la majorité des enquêtés (68.38%) déclarent qu'ils choisissent le «livre» dans leurs recherches documentaires, 12.82% choisissent les «Thèses et mémoires», 9% choisissent «autre», 8% indiquent «articles», mais seulement 3% choisissent "Périodiques".

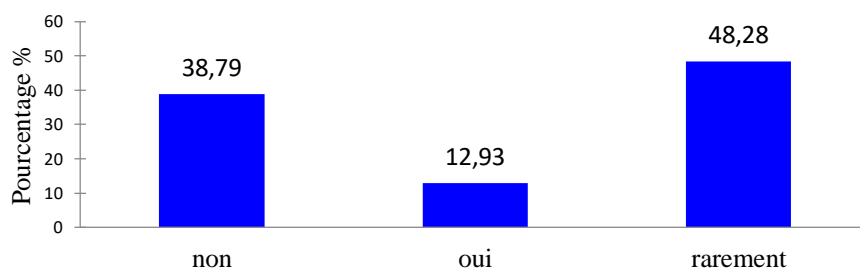


Figure 10 Obtention des documents demandés

48% des enquêtés trouvent rarement les documents demandés dans la bibliothèque ; 39% ne les trouvent pas, alors que seulement 5% trouvent l'information dont ils ont besoin. Ce résultat permet de conclure qu'il fallait résoudre cette problématique en pensant à des méthodes ou des outils informatiques qui permettent d'offrir la meilleure information au bon moment au bon lecteur (cf. figure 10).

Selon l'avis sur le projet d'informatisation de la bibliothèque, les résultats démontrent que 88% des enquêtés indiquent l'importance d'informatiser la bibliothèque, alors que seulement 12% déclarent qu'il n'est pas nécessaire.

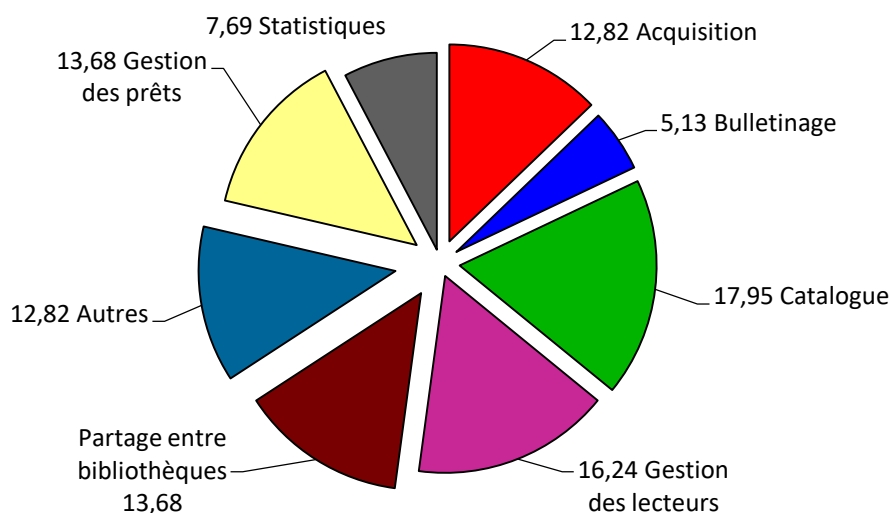


Figure 11 Principales fonctionnalités souhaitées à informatiser plus rapidement

D'après la figure 11, on remarque que les enquêtés ont des besoins diversifiés en terme des fonctionnalités souhaitées à informatiser au plus vite (entre 13% et 17%). Tandis que moins de 7 % d'eux souhaitent informatiser les outils statistiques et bulletinage.

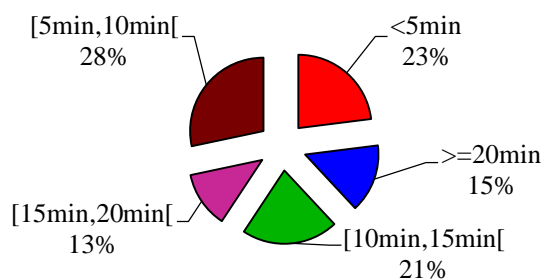


Figure 12 Temps passé à prêter un livre

Cette question vise à étudier l'efficacité des services en termes de temps passé à prêter un exemplaire. La figure 12 montre le lentement et la non fiabilité des services rendus par la bibliothèque traditionnelle.

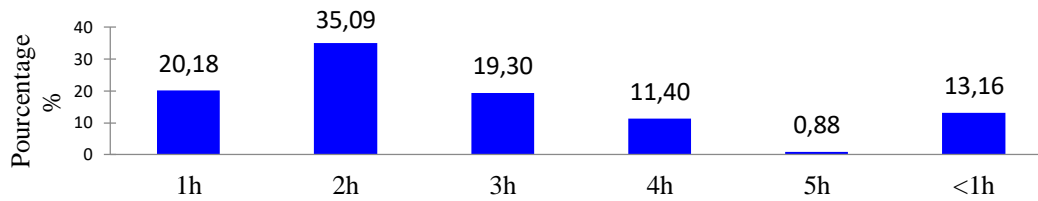


Figure 13 Temps passé à la bibliothèque

Le but de cette question est de savoir combien de temps les étudiants passent dans la bibliothèque (cf. figure 13). 35% des enquêtés passent 2 heures, 20% choisissent 1 heure, 19% indiquent qu'ils passent 3 heures, 13% mentionnent qu'ils passent moins d'une heure, mais seulement 11% choisissent 4 heures et 1% mentionnent 5 heures.

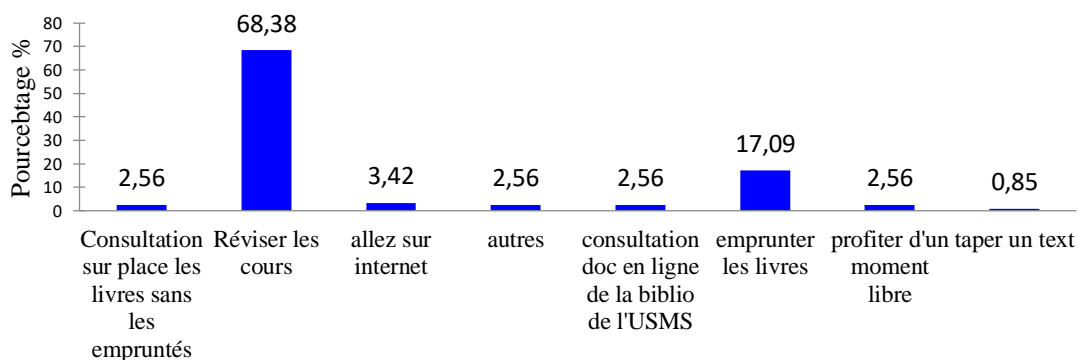


Figure 14 Raisons de visiter la bibliothèque

L'étude des motifs de visite à la bibliothèque nous renvoie à l'importance du lieu (lieu de travail). Pour 68% usagers, le premier motif de fréquentation de la bibliothèque universitaire est le travail sur place en révisant les cours. Ensuite, arrive l'emprunt des ouvrages (environ de 17 %) tandis que les autres motifs sont au moins de 3% (cf. figure 14).

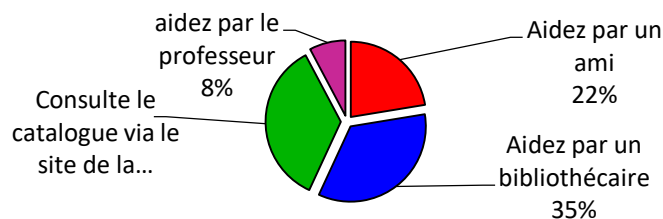


Figure 15 Aide principale pour rechercher des documents

Pour trouver le document ou l'information qu'ils recherchent, 35 % des usagers vont consulter le catalogue, 35% des étudiants se font aider par un bibliothécaire, 22% disent qu'ils reçoivent de l'aide d'un ami, alors que seulement 8% ils reçoivent de l'aide du professeur (cf. figure 15).

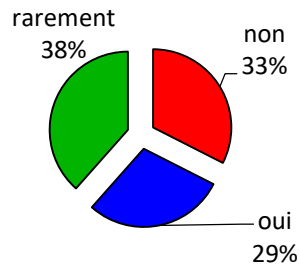


Figure 16 Enquêtés et ouvrages recommandés par le professeur

38% des enquêtés indiquent qu'ils choisissent rarement des ouvrages conseillés par le professeur, 32% des enquêtés ne les choisissent jamais, tandis que 29% déclarent les utilisent.(cf. figure 16)

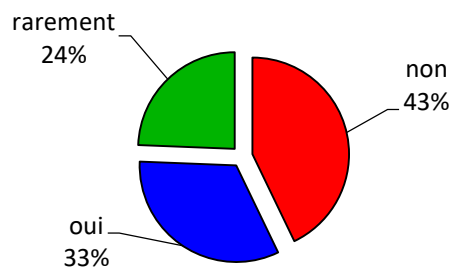


Figure 17 Enquêtés et l'habitude d'acheter des documents

La figure 17 indique que 43% des enquêtés n'achètent pas de documents, 33% ont l'habitude d'acheter des documents, tandis que 24% déclarent qu'ils n'achètent que rarement des documents.

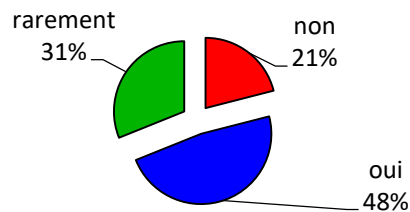


Figure 18 Lecture des documents d'autres disciplines

Environ la moitié des enquêtés (48%) lisent habituellement des documents provenant d'autres disciplines, tandis que 31% choisissent «rarement» et seulement 21% disent ne pas le lire (cf. figure 18).

De la figure 14 à 18, nous pouvons conclure que les étudiants de l'USMS ont besoin d'un outil automatique qui leurs permettent d'avoir accès à la bonne information au bon moment. Les institutions universitaires devraient inculquer à leurs étudiants les habitudes d'achat et leur permettre de lire d'autres documents pour d'autres domaines.

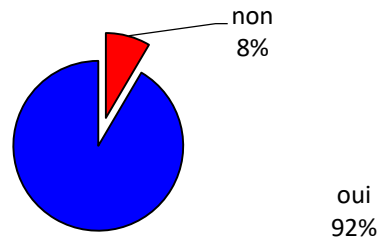


Figure 19 Possession d'un ordinateur personnel

La majorité des enquêtés (92%) déclarent qu'ils possèdent un ordinateur personnel, tandis que 8% disent ne le possèdent pas (Cf. figure 19).

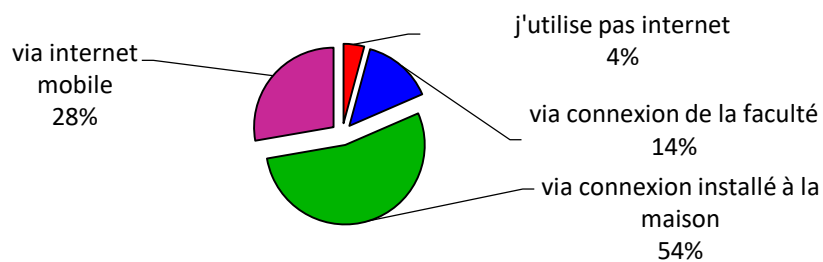


Figure 20 Méthode préférée pour accéder à Internet

Le but de cette question est de connaître la méthode préférée pour utiliser Internet, la moitié des enquêtés 54% préfèrent utiliser la connexion à la maison, 28% utilisent l'Internet sur un mobile, 14% utilisent la connexion de la Faculté, tandis que 4% disent que ils ne se connectent pas à l'internet (Cf. figure 20).

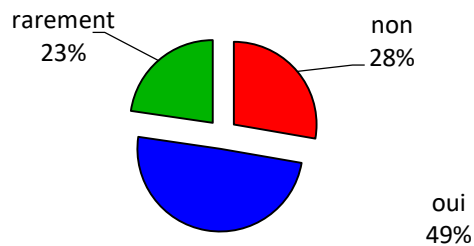


Figure 21 Utilisation de la technologie de l'information et de la communication

Environ de la moitié des personnes interrogées (49%) aiment communiquer avec le professeur en utilisant les nouvelles technologies de l'information et de la communication (TIC), tandis que 28% choisissent «non» et 22% seulement l'utilisent rarement (Cf. figure 21).

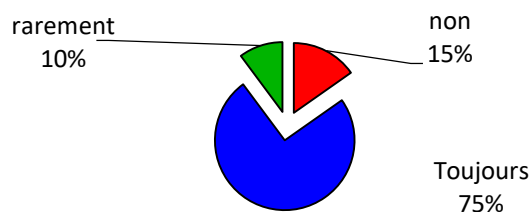


Figure 22 Utilisation de la documentation électronique

La figure 22 exprime que la majorité des enquêtés (75%) préfèrent utiliser la documentation électronique, tandis que 15% choisissent «Non» et seulement 10% disent qu'ils ne préfèrent que rarement.

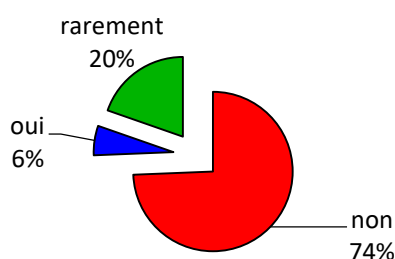


Figure 23 Activités à la bibliothèque

La majorité des enquêtés 74% indiquent qu'ils n'ont aucune activité dans la bibliothèque, tandis que 20% choisissent «rarement» et seulement 6% disent que la bibliothèque organise des activités (Cf. figure 23).

À partir du tableau indiqué dans l'annexe 3, nous concluons les résultats suivants :

- La moitié des enquêtés (55%) sont pour la plupart satisfaits des heures d'ouverture de la bibliothèque ; il appartient principalement à la Faculté de FP (59%), 21% sélectionnent «légèrement satisfait», tandis que 13% disent très satisfaits et seulement 12% choisissent «Pas satisfait».
- 43% des enquêtés sont pour la plupart satisfaits de la réception du bibliothécaire, il appartient principalement à la Faculté de FP (61%), 17% choisissent «légèrement satisfait», tandis que 33% se disent très satisfaits, tandis que 8% choisissent «Pas satisfait».
- La majorité des enquêtés (62%) sont légèrement satisfaits de l'état des locaux de la bibliothèque, il appartient principalement à la Faculté Polydisciplinaire (59%), 21% choisissent «satisfait», tandis que 16% se disent pas satisfait et seulement 1% choisissent «très satisfait».

b) Analyse descriptive (Analyse de la moyenne)

Dans cette analyse descriptive, nous utilisons une échelle à quatre points de 1 (pas satisfait) à 4 (très satisfait).

Tableau 1 présente la moyenne de la satisfaction des utilisateurs de trois services dans les bibliothèques en utilisant la moyenne des enquêtés.

Tableau 1: Analyse descriptive (moyenne)

Services	Moyenne des réponses
État des locaux	2.08
Réception du staff	3.01
l'horaire d'ouverture de la bibliothèque	2.68

D'après le tableau ci-dessus, on remarque qu'il y a une différence en comparant la moyenne des réponses, mais on ne peut pas juger qu'elle s'agit d'une différence fiable. Pour cela on va adopter d'autres techniques d'analyse de données telles que l'analyse de Pareto, l'ACM et la classification automatique des enquêtés.

c) Classification automatique non supervisé

i. Classification ascendante hiérarchique (CAH)

L'objectif de cette méthode est de classer tous les enquêtés dans les classes les plus similaires que possible. CAH est une méthode d'analyse multivariée qui permet de créer des groupes homogènes en utilisant la distance Euclidienne.

Tableau 2: Nombre des enquêtés dans chaque classe en utilisant la CAH

Numéro de classes	Somme	Pourcentage %	État de satisfaction
2	90	9	Pas satisfait
4	204	21	Moyennement satisfait
1	229	23	Satisfait
3	457	47	Très satisfait

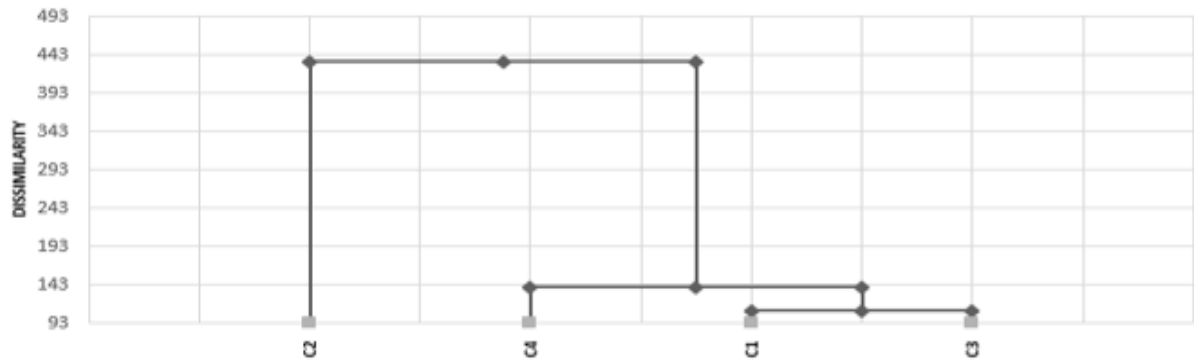


Figure 24 Dendrogramme de classes des enquêtés

La méthode CAH indique que 47% de réponses (Classe numéro 3) sont très satisfaits. 23% des enquêtés sont satisfaits et 21% sont légèrement satisfaits (classe 4), mais seulement 9% d'entre eux ne sont pas satisfaits de la majorité du service de bibliothèque. (Tableau 2)

ii. *KMeans*

Tableau 3: Nombre des enquêtés dans chaque classe en utilisant la méthode *KMeans*

Numéro de classe	total	Pourcentage %	État de satisfaction
2	90	9	Pas satisfait
4	220	23	Moyennement satisfait
1	343	35	Satisfait
3	327	33	Très satisfait

La méthode *KMeans* montre que les deux classes 1 et 3 des enquêtés sont satisfaits avec un pourcentage d'environ 33%, et seulement 9% d'entre eux ne sont pas satisfaits.

Les méthodes *KMeans* et CAH de la classification automatique permet de dépasser les limites de l'approche de l'analyse fréquentielle de l'enquête. Elles ont prouvé que le taux de satisfaction des services de la bibliothèque traditionnelle est encore faible.

d) **Analyse avec ACM**

On remarque que les personnes qui ont reçu une formation à la recherche documentaire, et ayant des connaissances informatique trouvent plus facilement les documents qu'elles recherchent (en bleu). (cf. figure 25)

La projection des données se faites selon les deux premiers axes (F1 et F2) qui présente 82,18% de l'information totale.

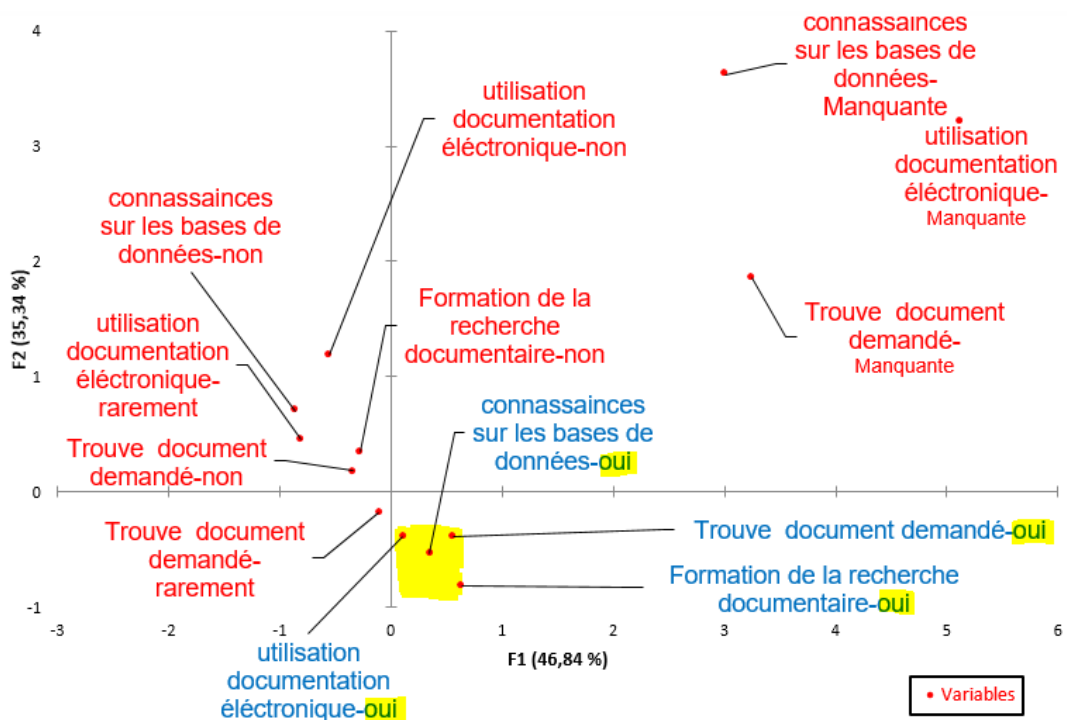


Figure 25 Étude de la relation formation, connaissances informatiques dans la recherche du document souhaité

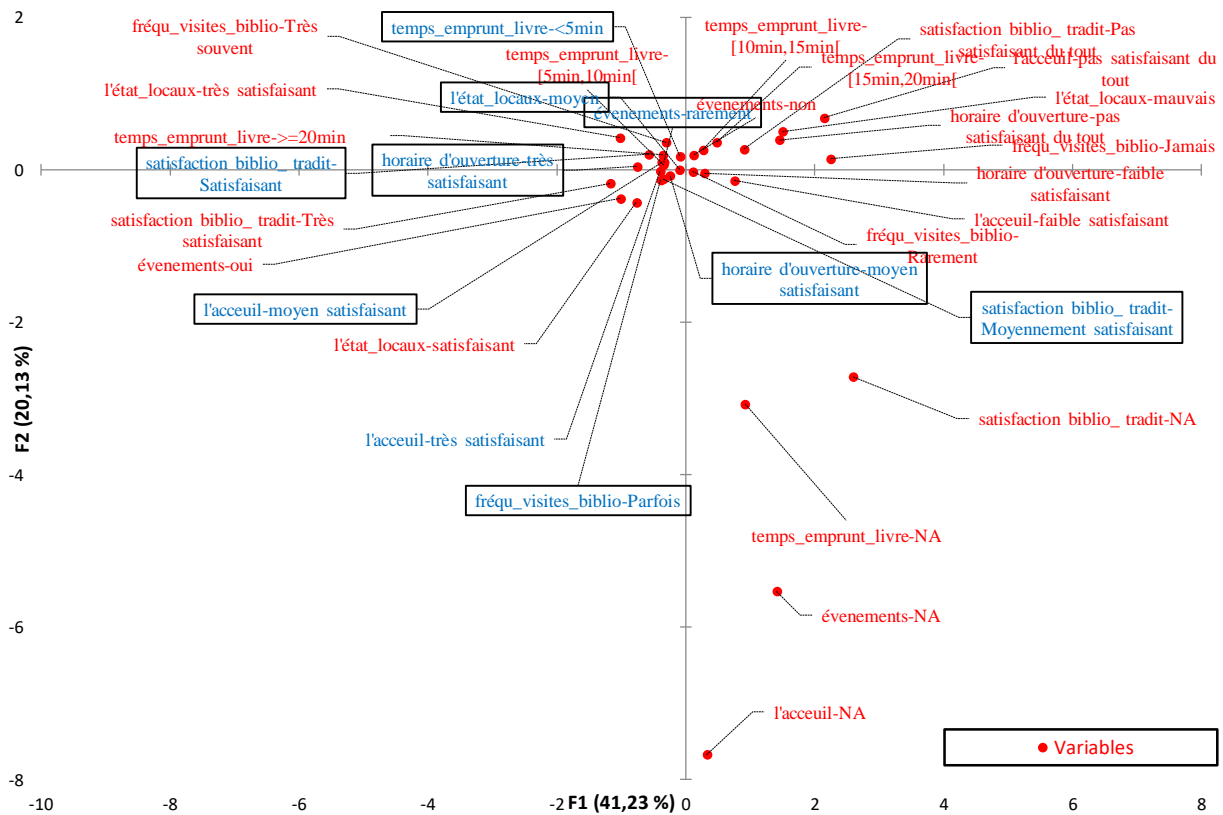


Figure 26 Résultat de l'enquête avec ACM

La figure 26 confirme les résultats retournés par les méthodes de classification automatique. En effet, les enquêtés sont légèrement satisfaits de toutes les services présentées par la bibliothèque (variables en bleu sont projeté dans la même zone du graphique).

e) Résultats de l'enquête en utilisant le nuage de mots

L'utilisation de nuages de mots (cf. figure 27) nous a permis de conclure que :

- Les étudiants sont moyennement satisfaits des services des bibliothèques traditionnelles. (cf. figure 27)
- La majorité des étudiants adopte l'aide du bibliothécaire pour chercher le document souhaité, ils confirment aussi que le projet d'informatisation de la bibliothèque est important.
- L'objectif principal de visiter la bibliothèque est de réviser les cours.

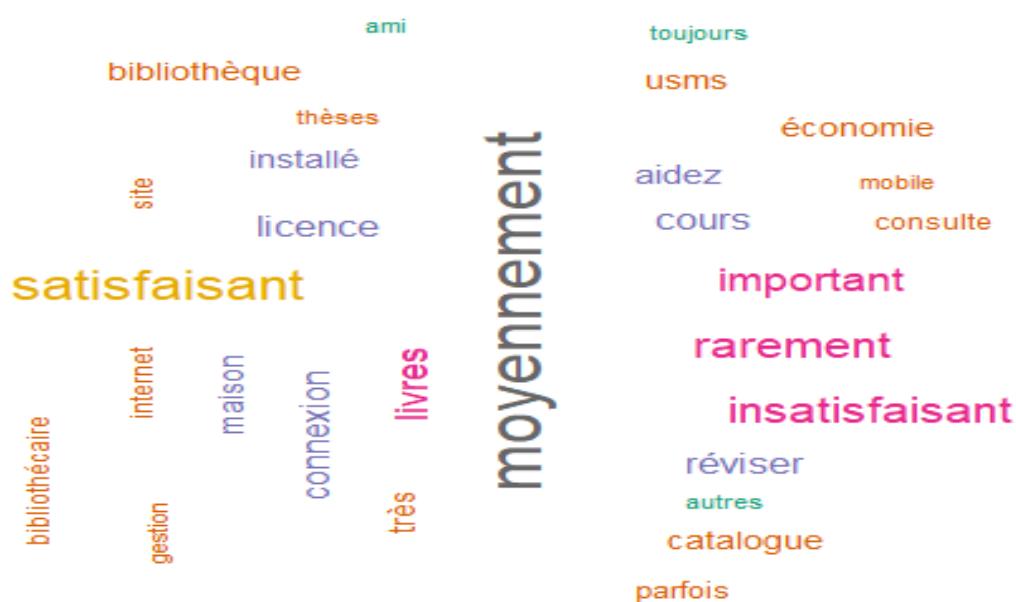


Figure 27 Nuage de mots extrait de l'analyse de l'enquête

D'après cette analyse, nous avons constaté qu'il est nécessaire d'instaurer un système numérique de gestion de bibliothèque qui comporte un système de recommandations aux étudiants.

f) Résultat de l'enquête en utilisant le diagramme de Pareto

Un diagramme de Pareto (loi 20/80 ou méthode ABC) est un outil analytique graphique permettant d'identifier les principales causes d'occurrence d'un phénomène en vue de mieux cibler les actions à mettre en œuvre ; il est considéré comme un outil pour prendre des décisions.

Le graphique de Pareto présenté dans la figure 28 est l'un des sept outils de base du contrôle de qualité. Cet outil sert à la priorisation des problèmes par ordre d'importance. La Loi de Pareto affirme que 20% des causes provoques 80% des effets.

Le tableau 4 présenté ci-dessous montre le résultat de l'analyse de Pareto.

Tableau 4 Résultat de l'analyse avec la loi de Pareto

Critère	Nombre des réponses négatives	Fréquence en%	cumulé en %
Activités dans la bibliothèque	898	11,17	12%
Obtenir le document demandé	825	10,27	24%
Satisfaction aux services de la bibliothèque traditionnelle	768	9,56	34%
Emprunte des documents recommandés par le professeur	678	8,44	43%
Habitude d'achat des documents	572	7,12	51%
Lecture de documents d'autres disciplines	506	6,30	58%
Utilisation des nouvelles TIC	490	6,10	65%
Temps passé à emprunter un document	449	5,59	71%
Fréquence des visites à la bibliothèque	392	4,88	76%
Utilisation d'un SGBD	335	4,17	81%
Temps passé à la bibliothèque	319	3,97	85%
Heures d'ouverture	319	3,97	90%
Utilisation de la documentation électronique	245	3,05	93%
Réception du staff	237	2,95	96%
État des locaux	155	1,93	98%
Avoir un ordinateur personnel	82	1,02	99%
Utilisation d'Internet	41	0,51	100%
Total	8036	100,00	

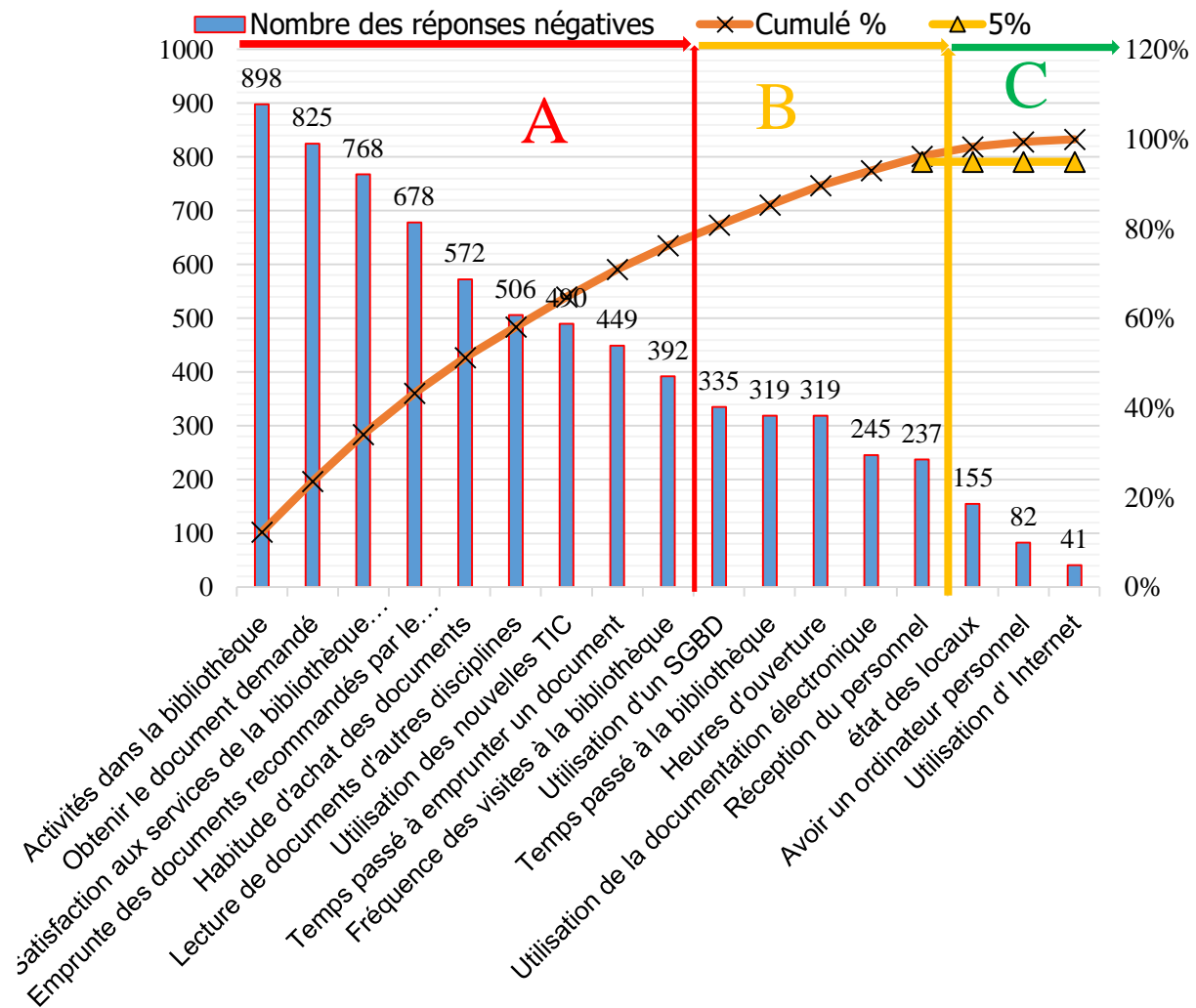


Figure 28 Diagramme de Pareto (analyse prioritaire)

Les problèmes à plus forts impacts se situent à gauche du graphique dans la région A (représente 80%), tandis que les moyens critiques dans la région B et finalement les plus tolérables dans la région C.

Le diagramme de Pareto, nous a permis de déduire qu'il faut prioriser les problèmes cités dans la région A (cf. figure 28) :

- Développer un outil pour faciliter la recherche de la meilleure information au meilleur moment pour la personne appropriée (obtenir le document demandé) ;
- Créer un espace approprié pour les étudiants de la bibliothèque (Organisation des activités) ;

- Réduire le temps passé par les étudiants à emprunter des documents (Temps passé à emprunter un document).
- Donner la priorité à l'informatisation de la bibliothèque (satisfaction de la bibliothèque traditionnelle)
- Encourager et former les étudiants à utiliser les nouvelles TIC, en particulier l'utilisation du catalogue OPAC (Online Public Access Catalog);
- Créer un espace approprié pour étudier dans les bibliothèques ;
- Guider les étudiants vers les documents les plus importants au bon moment ;
- Encourager l'emprunt et la lecture de documents d'autres disciplines.

1.5 Planification et budget prévisionnel du projet

Dans cette section, on abordera essentiellement le chronogramme d'exécution, un planning des acteurs et le budget prévisionnel du projet.

Pour cela, on exposera la démarche à suivre par le projet de mise en place des systèmes intégrés en faveur l'USMS. On commencera par la phase de définition du projet, ensuite la planification, puis l'exécution et on finira par la phase de la clôture. Avant de commencer ce projet, il sera nécessaire de tenir en considération ces deux outils suivants :

- Un projet est un point optimal qui doit respecter le trio qualité coût délai ;
- Les objectifs d'un projet doivent être de type SMART (Spécifique, Mesurable, Accepté, Réaliste et défini dans le temps).

La figure 29 suivante illustre le cycle de vie d'un tel projet par niveau d'effort :



Figure 29 Cycle de vie d'un projet

1.5.1 Phases de réalisation du projet

Dans le cas des petits projets, on peut passer à la phase de montage. Dans cette phase, il est nécessaire d'établir un cahier de charge fonctionnel ou charte de projet.

a) Phase de définition et du montage

Dans cette phase, on va cumuler des informations à propos du projet, pour cela on va construire une fiche de définition du projet en répondant aux questions suivantes :

- Qu'ils sont les enjeux et le contexte du projet ?

En premier lieu, l'objectif de cette question est de déterminer le besoin et le Maître d'ouvrage (MOA) du projet et en deuxième lieu, on cherche l'historique, les projets précédents, une analyse forces faiblesse SWOT, le périmètre et les contraintes du projet à gérer.

Pour répondre à la question précédente, nous avons adopté une étude de recensement des différents systèmes intègres de gestion au sein des établissements de l'USMS et une étude qualitative et quantitative d'enquête de satisfactions des lecteurs des bibliothèques de l'USMS ainsi que des entretiens auprès des responsables des établissements relevant à l'USMS.

Quels sont les résultats attendus ? Est-ce qu'un produit ou un service ? L'objectif est bien SMART ? Quelles sont les critères et les indicateurs mesurables de la réussite ou non du projet ?

Pour commencer le projet, il est nécessaire d'identifier quelques risques et de déterminer les scénarios défavorables qui feraient échouer ce projet. On peut utiliser les méthodes de résolutions des problèmes à savoir la loi Pareto, l'analyse SWOT (cf. figure 30) et via le diagramme Ishikawa [2] (cf. figure 31).

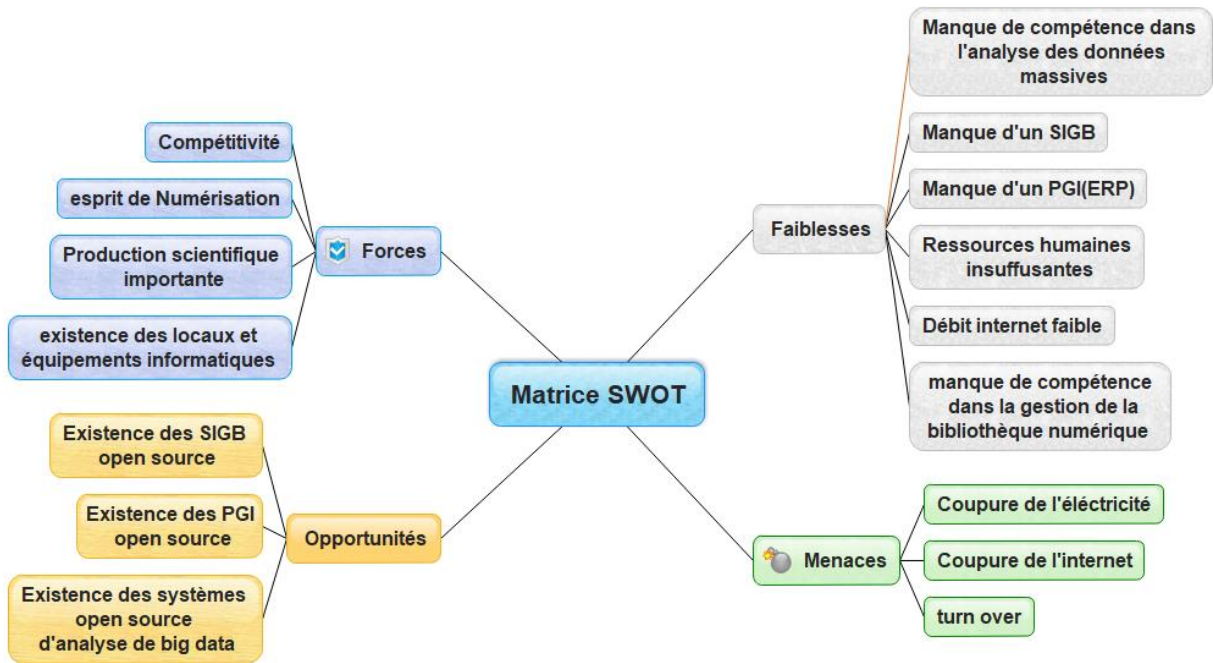


Figure 30 Analyse SWOT

Nous avons utilisé le Diagramme d'Ishikawa (Diagramme Cause effets, Fishbone diagramme, diagramme 5M/6M ou 7M) pour remonter les causes possibles des problèmes de l'insuffisance des systèmes intégrés de gestion de l'université.

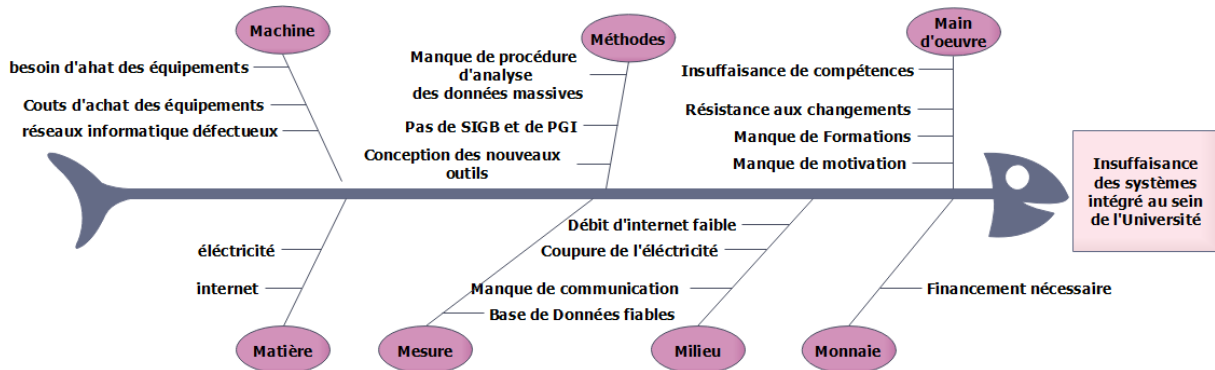


Figure 31 Diagramme 7M ou d'Ishikawa

- Quelles sont les moyens et ressources ?

Les moyens humains, financières, matériels et le temps nécessaire à la réalisation du projet ainsi que le besoin en expertise ...

- Quelles sont les acteurs (Membres de l'équipe) ?

Équipe, chef de projet (encadrant), client (faculté ou université).

La figure 32 illustre les caractéristiques du projet :

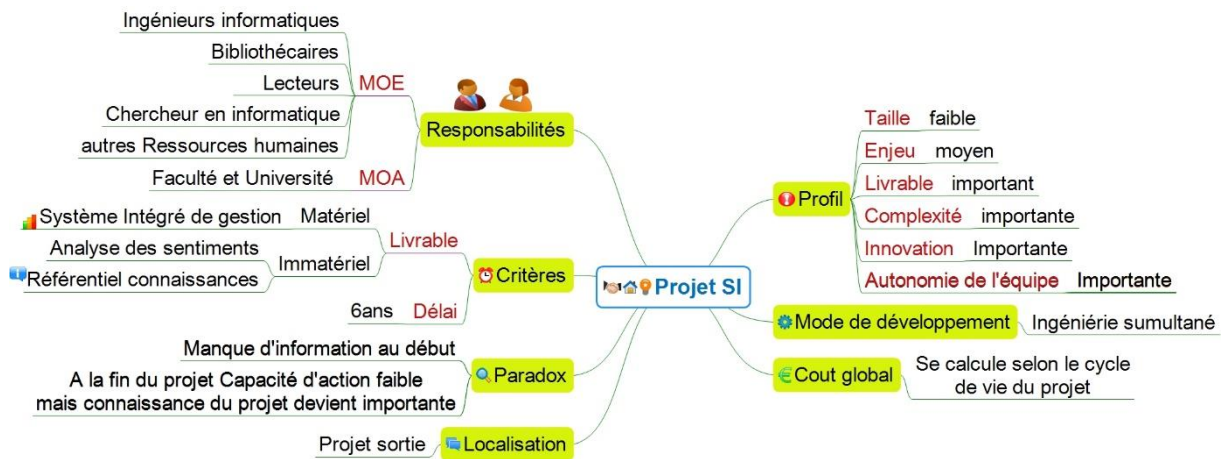


Figure 32 Caractéristiques du projet

A ce stade, nous avons défini le cahier de charge du projet ainsi nous avons réalisé une fiche de définition du projet, maintenant, on va entamer la partie de la planification du projet en vue de déterminer les tâches (WBS), les responsabilités (OBS), la durée ainsi que le coût global du projet.

b) Planification en utilisant le réseau PERT et Gantt

Avant de planifier un tel projet, il fallait définir les tâches (Works Breakdown Structure WBS) ainsi que leurs responsables (en anglais Organisationnel Breakdown Structure OBS) en les regroupant dans un diagramme RACI :

R : responsable ou coordonnateur ;

A : Acteur, qui réalise le travail ;

C : qui doit être tenu consulté avant de démarrer la tâches ;

I : qui doit être tenu informé du résultat.

Nous avons utilisé la technique PERT (abréviation en anglais du Program Evaluation and Review Technique) qui est une technique d'évaluation de la durée et de l'ordre des tâches et d'examen de programmes et de projets. Cette technique a été élaborée par une marraine américaine en 1957, qui a pu réduire un programme de développement (appelé Polaris) de 7 ans à 4 ans.

Le tableau 5 suivant résume l'ensemble des tâches du projet, l'antériorité des tâches ainsi que les acteurs intervenants.

Tableau 5 Tableau des antériorités des tâches

N° Tâche	Nom de la tâche	Durée en mois	Prédé- cesseurs	Acteurs
0	Projet de thèse	71		
1	<p>Analyse de l'existant :</p> <ul style="list-style-type: none"> ❖ Fiche de définition du projet : le contexte du projet, analyse SWOT, Diagramme 7M, détermination des acteurs, vérifications que les résultats attendus sont de type SMART ❖ Réalisation et analyse d'enquête par plusieurs méthodes 	6		Chercheur informaticien, lecteurs, décideurs, bibliothécaires, professeurs
2	État de l'art TAL, analyse des sentiments	6	1	Chercheur informaticien
3	État de l'art outils de visualisation	4	2	Chercheur informaticien
4	État de l'art Big Data	6	3	Chercheur informaticien
5	Banchmarking des SIGB	6	8	Chercheur informaticien
6	Banchmarking des ERP	6	8	Chercheur informaticien
7	Benchamrking des outillages d'analyse des données massives	6	8	Chercheur informaticien
8	Planification	1	4	Chercheur informaticien
9	Vérification des moyens informatiques existant	2	6	Chercheur informaticien, Ingénieurs, décideurs, bibliothécaires
10	Achat des équipements	2	9	Service financière, service matériel, chercheur informaticien
11	Installation PMB en local	5	9	Chercheur informaticien
12	Réalisation des notices bibliographiques	2	11	Service scolarité, Bibliothécaires, étudiants stagiaires, chercheur informaticien
13	Réalisation de la base de lecteurs	2	11	Service scolarité, Bibliothécaires, chercheur informaticien
14	Test SIGB en local	1	13;12	Chercheur informaticien

N° Tâche	Nom de la tâche	Durée en mois	Prédé- cesseurs	Acteurs
15	Installation de PMB en serveur	2	14	Chercheur informaticien, Ingénieurs, décideurs
16	Installation de l'ERP en local	2	15	Chercheur informaticien
17	Test ERP	2	15	Chercheur informaticien
18	Test de la gestion électronique des documents dans l'ERP	1	17;16	Chercheur informaticien
19	Évaluation de l'activité de la bibliothèque sous PMB	26	14	Chercheur informaticien
20	Analyse d'opinion des étudiants	6	18	Chercheur informaticien
21	Application de Hadoop HDFS pour l'analyse des données massives de l'université	6	20	Chercheur informaticien
22	Rédaction de mémoire	12	21	Chercheur informaticien
23	Livraison de documentation de l'installation et paramétrage des outils informatiques utilisés	2	22	Chercheur informaticien
24	Formation MOA, redoublonnage de la responsabilité	1	23	Chercheur informaticien
25	Fin	0	24	

Par la suite, nous avons adopté le diagramme de Gantt (élaboré par un ingénieur américain Henry Gantt en 1910) pour une présentation visuelle suivant un calendrier du temps de l'ensemble des tâches du projet. (cf. figure 33)

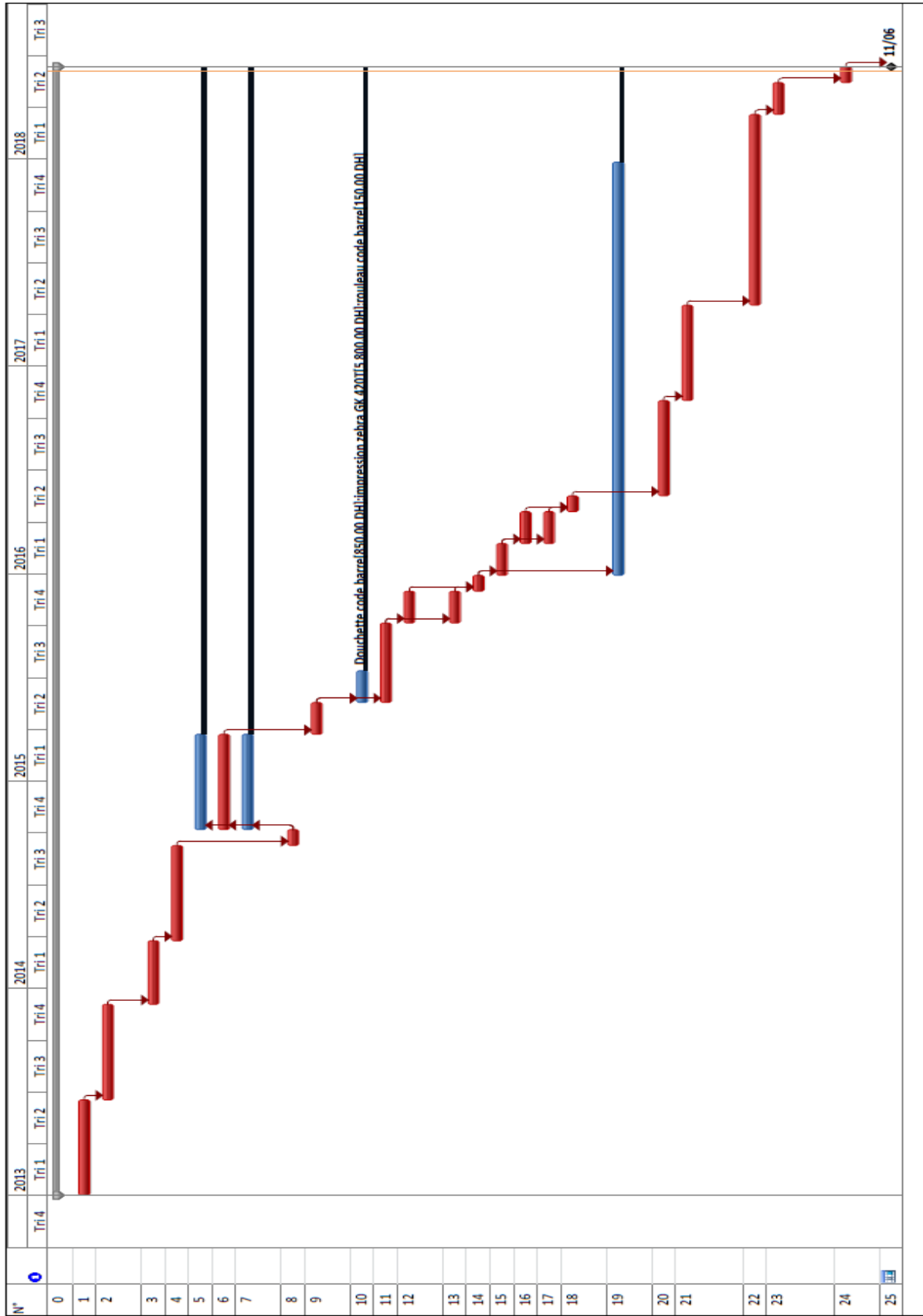


Figure 33 Diagramme de Gantt du projet

c) **Suivi et évaluation du projet**

Pour réaliser un bon suivi du projet, nous avons défini pour chaque tâche des contraintes à gérer (par exemple le problème financière, la disponibilité des bibliothécaires, de l'ingénieur informaticien, l'accès aux réseaux internet...), les résultats attendus (doivent être de type SMART), un échéancier ainsi que la distribution des acteurs.

Le coût global prévisionnel du projet est de 6800 DH, ce coût a été calculé à la base des coûts d'achat des équipements (douchettes, rouleau des codes à barres et imprimante zebra GK), et la fin du projet sera le 11/06/2018.

1.6 Conclusion

Dans ce chapitre, nous avons tout d'abord abordé le cadre institutionnel de l'étude, une analyse de l'existant (recensement des différents systèmes intégrés au sein de l'université et via une enquête de satisfaction autour d'un service). Nous avons constaté qu'il est nécessaire de penser à revoir la manière dont l'établissement gère leurs documents et les attentes des différents intervenants acteurs.

Ensuite nous avons traité la planification de la mise en œuvre de ce travail de recherche comme étant un projet informatique. En premier lieu, nous avons présenté d'une manière détaillée le cahier des charges qui fera la description détaillée du projet. En deuxième lieu, nous avons proposé une planification de la mise en œuvre du projet en évaluant le temps d'exécution à travers un chronogramme d'exécution des actions (GANTT) et un planning d'intervention des différents acteurs. Finalement, nous avons élaboré un plan d'action de suivi-évaluation et nous avons présenté le budget prévisionnel du projet.

Le chapitre suivant sera consacré à présenter l'état de l'art sur les systèmes d'information de la base de données relationnelles à l'analyse de données massives (Big data).

Chapitre 2. De la base de données relationnelle à l'analyse de données massives (Big Data)

Chapitre 2. De la base de données relationnelle à l'analyse de données massives (Big Data)

2.1 Introduction

Dans ce chapitre, on présentera l'évolution du SI de la base de données relationnelle (données structurées) au Big Data (données assez volumineuses structurées, semi ou non structurées) où nous donnerons les méthodes de conception des SI, les diverses techniques de collecte et d'analyse des données à savoir l'analyse descriptive, l'analyse exploratoire, les outils d'analyse d'opinions et les outils de visualisation de données.

2.2 Définition d'un Système d'Information

L'information devient aujourd'hui très indispensable pour toute entreprise ou organisme. Elle est définie d'après l'auteur [3] comme des données à l'état brut (données non compréhensibles) qui deviennent compréhensibles en les ajoutant le contexte (Figure 34). La connaissance est obtenue avec une bonne utilisation et compréhension de ces informations tant que la sagesse (Compétence) est obtenue avec la compréhension des principes qui régissent les modèles de cette connaissance.

La figure présentée ci-dessous indique l'évolution de l'information selon Balmissse [3]:

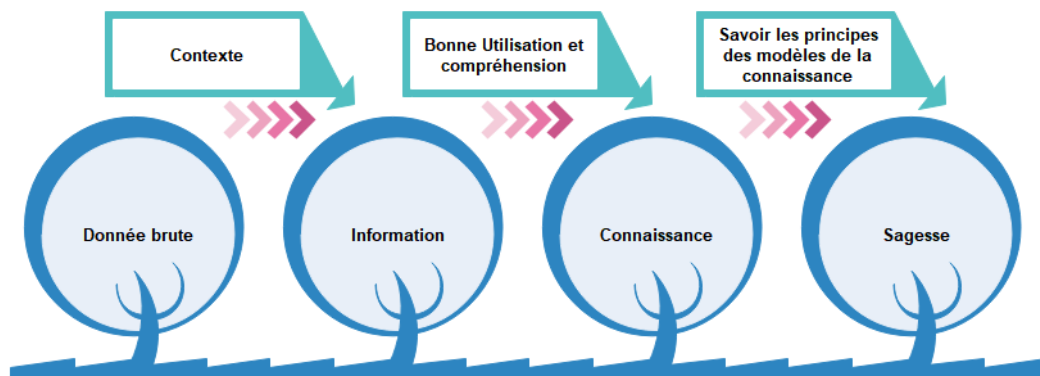


Figure 34: Définition de l'information selon Balmissse

D'après la littérature [4], un système d'information est défini comme un ensemble organisé de ressources (matériel, logiciel, personnel, données, procédures...) qui collectent, traitent, stockent les informations (sous forme de données, textes, images, sons, etc...) dans et entre les organisations.

La figure suivante présente la décomposition systématique de l'entreprise :

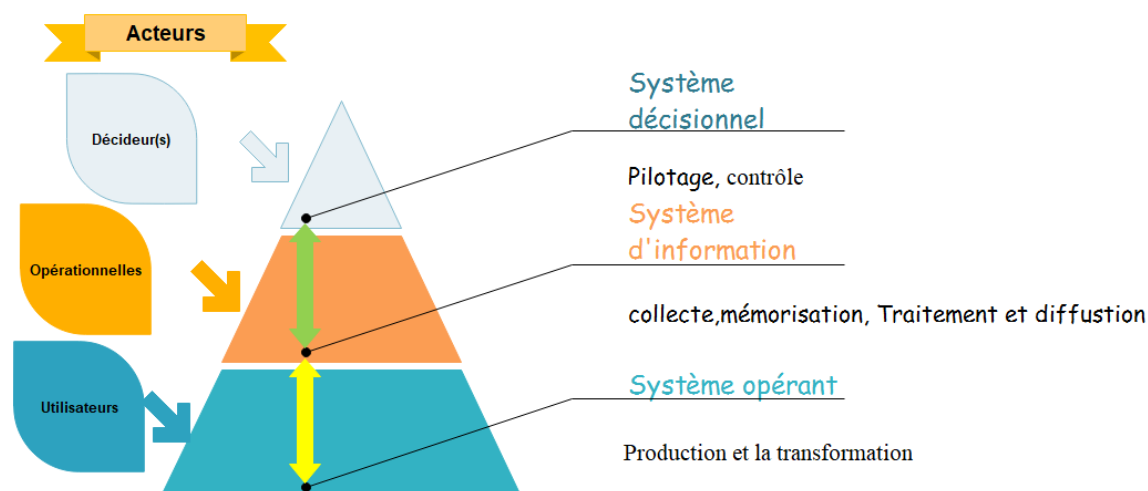


Figure 35 : Décomposition systématique de l'entreprise

- Le système opérant se compose de l'ensemble des ressources relatives à l'activité de l'entreprise.
- Le système décisionnel (de pilotage) assure la conduite des services de l'entreprise.
- Le système d'information assure l'interface entre le système opérant et le système de décision.

L'auteur [5] a défini le système informatique comme étant un ensemble de composantes informatiques inter-reliées qui recueillent de l'information, la stockent, la traitent et la diffusent afin d'aider à la prise de décision et au contrôle dans l'organisation.

Le Système Informatique a pour but de dépasser les limites des Systèmes d'Information traditionnelles (redondance et incohérence de données, manque de sécurité, de disponibilité et de partage de données...).

Ce qui permet d'assurer la qualité du Système Informatique :

- La confidentialité et la sécurité ;
- La fiabilité : mettre à jour en permanence de l'information ;
- La pertinence de l'information : avoir l'information adéquate ;
- Et la rapidité : obtenir une information rapide pour prendre des bonnes décisions dans des délais voulus.

On peut avoir plusieurs systèmes d'information :

- Le système d'information en ressources humaines (SIRH) est définie par Tannenbaum [6] comme un système permettant d'acquérir, de stocker, de manipuler, d'analyser, d'extraire et de distribuer des informations pertinentes au regard des ressources humaines d'une organisation.
- Le système d'information Géographique SIG : pour recueillir, stocker, traiter, analyser, gérer et présenter tous les types de données spatiales et géographiques.
- Système d'information pour les entreprises agricoles et agroalimentaires.
- Etc.

Pour modéliser un système d'information, on peut suivre les étapes habituelles présentées dans la figure ci-dessous :

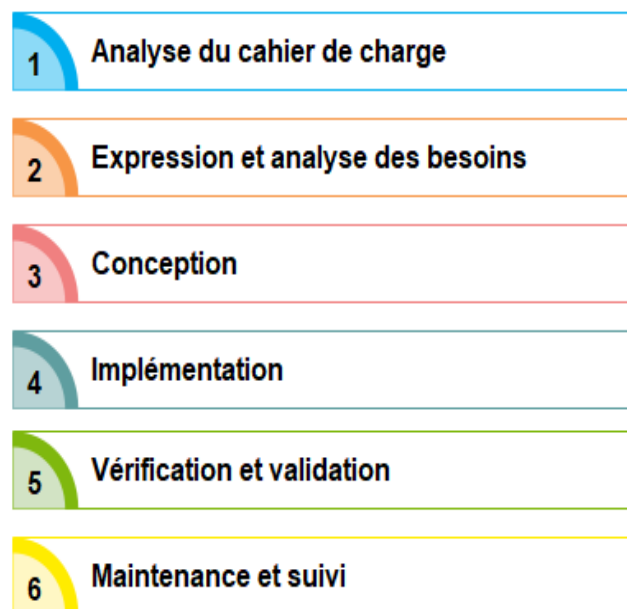


Figure 36 : Phases de réalisation d'un système d'information

2.3 Historique des SI dans les Universités Marocaines

Les SI sont apparus dans les années 1970. L'entreprise n'y pas considérée comme un ensemble de services indépendantes mais un ensemble d'éléments, de moyens humaines et techniques en relation qui est ouvert sur l'environnement. Les SI des Universités Marocaines connaissent une croissance quasi exponentielle des données et des informations qui ne sont pas homogènes. Alors que les décideurs au niveau central ne peuvent pas d'en sortir des indicateurs précis et fiables pour prendre des meilleures décisions dans le bon moment. La majorité des

universités nationales ont opté pour Apogée, d'autres continuent de travailler toujours avec leurs applications propres payantes, ou font appel à des progiciels commerciaux (Microsoft office, SPSS, SPHINX...) ou d'autres applications spécifiques à certaines module de gestion (gestion de congé, gestion d'absence, gestion de paiement, gestion commercial...).

2.4 Méthodes de conception d'un système d'information

La conception d'un système d'information fait recours à des modèles : le modèle E/A (Entité association) élaboré par Mr Peter Chen en 1976 [7], puis l'apparition d'une Méthode d' Étude et de Réalisation Informatique pour les Systèmes d'Entreprise(MERISE) en 1978 à la France, ensuite il vient le langage de modélisation unifié UML(en anglais Unified Modeling Language) en 1997, et enfin MERISE a été étendu en MERISE2.

2.5 Système de gestion de Base de données relationnelle et langage SQL

2.5.1 Base de Données Relationnelle

Les bases de données (BD) connaissent une évolution au niveau de la manière dont les données sont structurées passant du modèle hiérarchique, puis en réseau en arrivant au modèle relationnel.

Une BD est un ensemble structuré de données décrites sous forme de modèle de données enregistrées sur des supports informatiques (CD, Disque durs...) accessibles de manière confidentielle par plusieurs utilisateurs (privileges d'accès contrôlé).

La BD relationnelle est gérée par un progiciel dénommé le Système de Gestion de Base de Données Relationnelle (SGBDR).

2.5.2 SGBDR

Système de Gestion de Base de Données Relationnelle (SGBDR, en anglais Relational Data Base Management System) est un progiciel qui supporte un modèle de données qui permet de décrire, de manipuler et d'assurer la sécurité et la confidentialité des données dans un environnement où de nombreuses utilisateurs ayant le besoins variés peuvent interagir simultanément sur les données. Ce progiciel a pour objectif de contrôler l'accès, la redondance et l'incohérence et de faciliter la manipulation de données. Ces grands principes sont la

combinaison de clefs primaires et étrangères pour établir un lien entre tables, liens remontés par le biais de requêtes SQL. (Voir Annexe 1)

2.5.3 SQL

On utilise le langage de requêtes SQL (Structured Query Language) pour interroger les bases de données relationnelles. Il est introduit par IBM, c'est une évolution du langage SEQUEL. Il est devenu le langage standard pour décrire et manipuler les Bases de Données Relationnelles.

L'évolution d'espace de stockage, la production de l'information et le besoin d'y accéder de manière plus rapide et pertinente mènent à l'apparition des nouvelles modèles d'analyse de données parmi lesquels le modèle multidimensionnel ainsi par la suite les outils d'analyse de données massives (Big Data).

2.6 Modèle multidimensionnel

Le modèle multidimensionnel permet d'analyser les données sur plusieurs axes (en général 3 axes) en formant ainsi ce qu'on appelle des hyper cubes par contre le modèle relationnelle n'a que deux axes.

Pour utiliser l'analyse multidimensionnelle des activités des entreprises, il faut les modéliser suivant deux types des axes (axe de fait et axe de dimension) en répondant à ces deux questions :

- On veut mesurer quoi ? c'est axe de fait.
- Sur quelles dimensions ? sont les axes de dimension.

Par exemple, on cherche à aboutir le nombre de vente pour un produit donné par client. Pour illustrer ceci, la figure 37 présente le modèle en étoile de la conception (axes de dimensions : Client, produit, date, région et axe de fait central : vente) ainsi que la vue multidimensionnelle des 3 axes produit, vente et client.

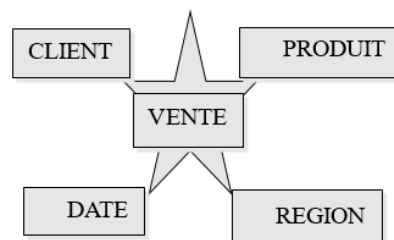


Figure 37 Modèle multidimensionnel en étoile

L'OLAP (On-line Analytical Processing) est une technique d'analyse de données multidimensionnelles élaborée pour la première fois par le créateur de base de données relationnelle E.F.Codd en 1993, cet outil informatique permet aux décideurs des entreprises, selon leurs besoins, d'accéder rapidement et interactivement à une information pertinente présentée selon plusieurs axes d'analyse. [8]

2.7 Architecture et fonctionnement d'un système décisionnel

2.7.1 Entrepôt de données et base métiers

Selon l'auteur [9], un entrepôt de données est une collection de données orientées sujet, intégrées, non volatiles, historisées et organisées pour le support d'un processus d'aide à la décision.

On peut subdiviser l'architecture d'un système décisionnel en quatre niveaux comme indiqué dans la figure 38 :

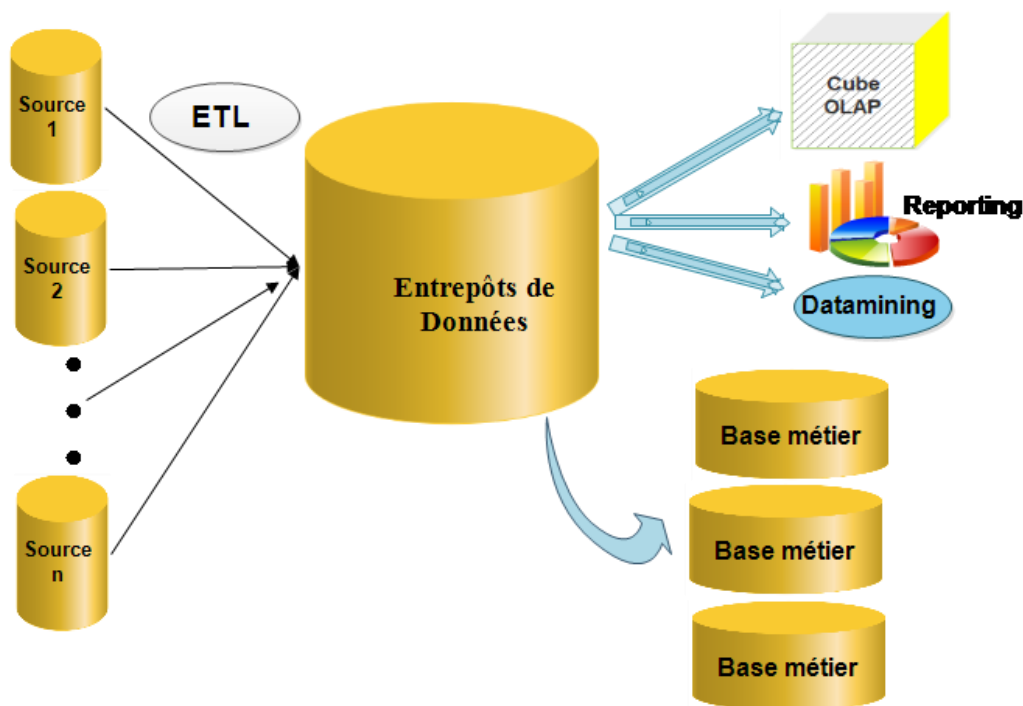


Figure 38 Architecture d'un Système d'Information décisionnel

- Le niveau 1 : concerne les différents sources d'information (fichier plats, bases de données Access, BD Microsoft SQL server, BD oracle, fichier Excel, csv, et autres...)

- Le niveau 2 : on subit aux différentes sources du niveau 1 quelques modifications, standardisations après avoir l'extraire en utilisant l'outil ETL (extraction, transformation et chargement) puis implémentation d'un datawarehouse.

Il existe plusieurs ETL open source à savoir Kettle (Pentaho data integration), Talend open studio, CloverETL...

- Le niveau 3 : c'est l'entrepôt de données elle-même qui peut être traité par métier (datamart).
- Le niveau 4 : l'utilisateur final peut construire sa propre question décisionnelle en traitant le datawarehouse via l'utilisation de reporting, par le biais de l'OLAP ou par le forage de données (datamining).

Une base métier ou magasin de données (datamart) est un sous-ensemble d'un datawarehouse par métier destiné à fournir des données aux utilisateurs en vue de faciliter la prise de décision.

2.7.2 Outils de reporting

Les outils de Reporting répondent au besoin principal des utilisateurs, à savoir : produire des rapports et des tableaux de bords. Il existe de nombreux outils de reporting dans le marché tels que Jasper, JFreechart, powerBI, etc...

- Outil power BI

L'outil Power BI (Business Intelligence) est une suite d'outil proposé par Microsoft permettant de collecter, de rechercher, de transformer, de se connecter à diverses sources de données, d'analyser et de mettre en forme les données dans des tableaux de bord interactifs et collaboratifs.

- Jasper (ireport) : est un outil de reporting open source, offert sous forme d'une bibliothèque qui peut être embarquée dans tous types d'applications Java.
- JFreechart : est une API Java open source permettant de créer des graphiques et des diagrammes de meilleure qualité.

- BIRT (Business Intelligence and Reporting Tools): Un outil open source de la fondation Eclipse basé sur la technologie Java permet de produire des rapports graphiques synthétiques.

Dans notre étude, nous avons adopté le Power Designer BI comme outil de reporting.

Dans ce qui suit, on présente les outils d'analyse de données, de statistique et les méthodes d'exploration ou le forage de données (datamining).

2.8 Analyse de données, outils statistiques et forage de données (Data Mining)

2.8.1 Statistique descriptive

La statistique descriptive à deux dimensions a pour objectif de caractériser les relations qui existent entre deux séries d'observations qui peuvent être de nature :

- Qualitative : nominal ou ordinal (couleur des yeux, groupe sanguin...)
- Quantitative (quantifiable) : continue ou discontinue (le nombre de j'aime sur Facebook, le poids...)

La statistique descriptive sert à résumer les données brute grâce à des calculs (moyenne, écart type...) ou des présentations simples (Box plot, barre...)

Le box plot ou la Boîte à moustaches sert à présenter la tendance des données par rapport à la moyenne ainsi qu'elle donne une vue sur les points les plus extrêmes.

Pour analyser la tendance par rapport à la moyenne, on utilise la Variance et l'Écart type (la dispersion des données par rapport à la moyenne). La différence entre la médiane et la moyenne est que la première n'est pas sensible aux valeurs extrêmes. Le tableau 6 présente les différents types d'analyse selon le type de données. Le tableau suivant présente les types des graphiques et d'analyse par type de données :

Tableau 6 Choix d'une technique d'analyse de données descriptive

Type de données	Type d'analyse ou de graphique
1 variable quantitative	Moyenne, écart type, Tri à plat, Box plot
1 variable qualitative	Tri à plat, mode, camembert, Box plot
1 variable quantitative vs 1 var quantitative	Nuage de points
1 variable qualitative vs 1 var qualitative	Tri croisé (table de contingence)
2 variable quantitatives vs 1 var qualitative	Nuage de points + coloration par modalité de la variable qualitative
1 variable quantitative vs 1 variable qualitative	Boîte à moustaches
1 variable quantitative	Moyenne, écart type, Tri à plat, Box plot
1 variable qualitative	Tri à plat, mode, camembert, Box plot
1 variable quantitative vs 1 var quantitative	Nuage de points
1 variable qualitative vs 1 var qualitative	Tri croisé (table de contingence)
2 variable quantitatives vs 1 var qualitative	Nuage de points + coloration par modalité de la variable qualitative
1 variable quantitative vs 1 variable qualitative	Boîte à moustaches

2.8.2 Statistique exploratoire

Les méthodes exploratoires d'analyse de données multi-variées permettent souvent une réduction de dimension d'un jeu de données assez volumineux pour faciliter sa segmentation, d'étudier aussi la corrélation pour tester le lien entre deux groupes de variables.

Les méthodes de statistiques exploratoires sont subdivisées en 3 grandes catégories : méthodes de corrélation, de segmentation et de décisions.

a) Méthodes exploratoires de corrélation

1. Analyse en Composante Principale (ACP)

C'est une méthode exploratoire d'analyse de données quantitatives [10] quelles que soit leurs natures continues ou discrètes. Elle permet de visualiser et analyser rapidement les corrélations entre les N variables d'un tableau de jeux de données de M observations (cf. figure 39). L'ACP a pour objectif de simplifier un tableau de données initiales en passant d'un grand nombre de variables à un petit nombre de variables obtenues en regroupant les variables initiales présentant un coefficient de corrélation relativement important.

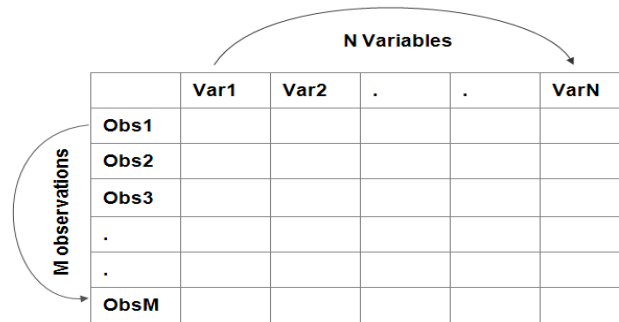


Figure 39 Jeux de données d'ACP

Les avantages de l'ACP :

- Permettre la projection des jeux de données assez grandes dans deux ou trois dimensions.
- Permettre l'étude de la corrélation entre plusieurs variables quantitatives en étudiant seulement la valeur du coefficient de corrélation R présenté dans la matrice de corrélation.
- Permettre l'aboutissement des cartes de représentation sur lesquelles on peut visuellement observer les proximités entre les catégories des variables qualitatives et les observations.
- Permettre la création des cartes de projection sur lesquelles on visualise les proximités entre les variables et les observations.

Soit deux séries $X(x_1, x_2 \dots x_n)$ et $Y(y_1, y_2 \dots y_n)$, x_i est la valeur de la variable pour l'individu i .

La moyenne simple se calcule par l'équation ci-dessous où N désigne le nombre d'individus.

$$\bar{X} = \frac{\sum_{i=1}^N x_i}{N} \quad (1)$$

Le coefficient de corrélation linéaire R s'écrit par la formule suivante :

$$R = \frac{\text{cov}(X,Y)}{\sigma_X \times \sigma_Y} \quad (2)$$

Avec σ_x est l'écart type de la série X ; σ_y est l'écart type de la série Y
 $\text{cov}(X, Y)$ désigne la covariance des séries X et Y.

Si R vaut 0, les variables ne sont pas corrélées. Mais elles sont d'autant mieux corrélées que R est loin de 0 (proche de -1 ou 1).

2. Analyse des Correspondances Multiples ACM

ACM [11] est une méthode qui permet d'étudier la corrélation entre au moins deux variables qualitatives ce que l'ACP ne traite que les variables quantitatives.

3. Analyse Factorielle des Correspondances (AFC) [12]

C'est un cas particulier de l'ACM. Il permet d'analyser seulement deux variables qualitatives.

4. Analyse Factorielle Multiple AFM

Elle a été élaborée par les auteurs [13], cette méthode permet d'analyser simultanément plusieurs tableaux de variables de différentes catégories (quantitatives et qualitatives), et d'obtenir des représentations graphiques, qui permettent d'étudier la relation entre les observations, les variables et les tableaux.

b) Méthodes exploratoires de segmentation par apprentissage non supervisé

Dans ce qui suit, on présente la liste des techniques de segmentations.

1. Kmeans [14]

K-Means clustering ou la nuée dynamique [14] est une méthode couramment utilisée pour regrouper automatiquement un ensemble de données en k classes. Elle procède en sélectionnant le centre initial des k cluster, puis les affiner itérativement comme indiqué dans la figure ci-dessous:

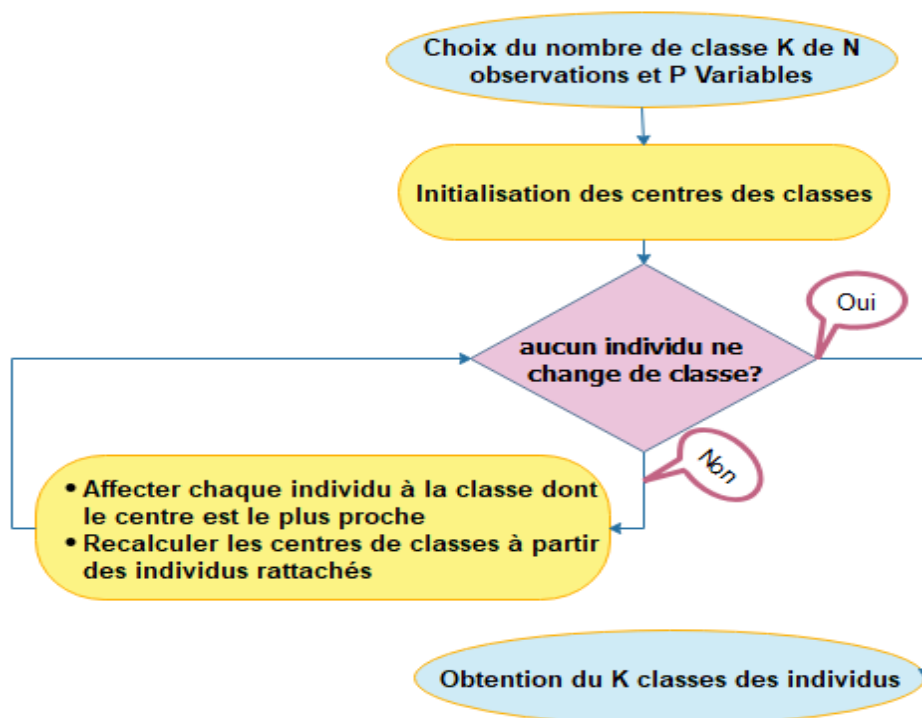


Figure 40 Algorithme de Kmean

Au début, on détermine le nombre de cluster K et on suppose le barycentre ou le centre de ces groupes. On peut prendre tous les objets aléatoires comme des centroides initiales ou les

premiers objets K en séquence peut également servir de centroides initiales. Puis on choisit la distance qu'on va utiliser afin de déterminer la distance minimale de chaque objet au barycentre. Dans cette application, nous avons utilisé la distance euclidienne qui est définie comme suit :

$$d(w_i, w_j) = \sqrt{\sum_{i=1}^n (x_i - x_j)^2} \quad (3)$$

avec $w_i = (x_1, x_2 \dots x_n)$ et $w_j = (y_1, y_2 \dots y_n)$

Ensuite, on détermine les coordonnées de barycentre, puis, on calcule la distance de chaque objet au barycentre (critère de dissimilarité). Enfin on groupe l'objet en fonction de la distance minimale (critère d'agrégation).

L'inconvénient de cette méthode est qu'elle permet toujours de choisir le nombre de classe à priori ce qui ne permet pas de découvrir qu'elle peut être un nombre cohérent de classes.

Nous avons utilisé cette méthode en vue de classifier les individus d'une enquête en les regroupant en des classes les plus homogènes que possibles. Les classes obtenues sont calculées à la base de la distance Euclidienne minimale (critère de dissimilarité et d'agrégation).

2. Classification ascendante hiérarchique CAH

La Classification Ascendante Hiérarchique [15] (ou "CAH") considère initialement toutes les observations comme étant des clusters ne contenant qu'une seule observation (singleton), et leur distance est alors le plus souvent définie comme étant leur distance euclidienne. La première étape consiste donc à réunir dans un cluster à deux observations les deux observations les plus proches. Puis la CAH continue, fusionnant à chaque étape les deux clusters les plus proches au sens de la distance choisie.

Le processus s'arrête quand les deux clusters restant fusionnent dans l'unique cluster contenant toutes les observations.

La réalisation de la classification ascendante hiérarchique se fait en trois étapes:

- *Étape 1 : Choix d'un indice de dissimilarité*

De nombreuses mesures de la distance [16] entre individus ont été proposées (Distance Euclidienne, distance de Manhattan, distance de Chebyshev, distance de Hamming...). Le choix d'une (ou plusieurs) d'entre elles dépend des données étudiées.

On dessine un tableau de distance qui contient les distances euclidiennes entre les individus de la population.

- *Étape 2 : Choix d'un indice d'agrégation*

On fait un regroupement selon la distance euclidienne minimale. Pour cela, on parcourt le tableau de distance pour obtenir le couple (x^*, y^*) , et on réunit deux éléments qui sont voisins dans une même classe $A=x^* \cup y^*$. Les autres classes sont inchangées.

- *Étape 3 : On calcule les distances euclidiennes entre la classe A et les autres points différents à A.*

Condition d'arrêt consiste à atteindre une seule classe comportant tous les éléments.

c) **Méthodes exploratoires de segmentation par apprentissage supervisé**

Ce sont des méthodes de classification supervisée dont on dispose d'une base de données d'apprentissage de K observations étiquetées pour estimer la sortie associée à une nouvelle observation x.

1. Machines à vecteur de support (SVM)

Machines à vecteur de support (Support Vector Machines) est une technique d'apprentissage supervisé. Elle permet de résoudre des problèmes de discrimination. On peut utiliser cette technique dans plusieurs domaines à savoir bio-informatique, finance, recherche d'information...

Cette technique aussi est largement utilisée dans la classification des textes [17] et précisément dans l'analyse des sentiments [18].

2. K plus proche voisin KNN(K-Nearest Neighbour) [19]

C'est une technique de classification supervisée qui permet d'estimer la sortie y associée à une nouvelle entrée x selon une distance à définir, en tenant compte les K échantillons d'apprentissage.

3. Réseau de neurones artificiel ou ANN (en anglais Artificial Neural Networks)

Le réseau de neurones artificiel ou ANN sont des constructions abstraites inspiré de l'activité d'un réseau de neurones biologique simplifié. Cette technique a été utilisée dans l'analyse des sentiments dans les travaux des auteurs [20] , [21] , [22] , [23] et [24].

4. Bayes naïf

La classification naïve bayésienne ou Bayes naïf [25] est un type de classification bayésienne probabiliste simple basée sur le théorème de Bayes avec une forte indépendance. On peut utiliser ce classifieur en vue de faire une classification et la catégorisation des documents.

Étant donnée la probabilité $P(w_i|C)$ d'appartenant un $i^{\text{ème}}$ mot (w_i) dans une classe C et $P(C)$ est la probabilité des classes des documents.

Puisque les probabilités sont indépendantes, donc nous avons :

$$P(D|C) = \prod_i P(w_i|C) \quad (4)$$

Donc la méthode bayes naïf permet de calculer la probabilité d'appartenance d'un nouveau document à une catégorie ou classe $P(C|D)$ sachant qu'on nous avons $P(C)$ et $P(C \cap D)$

2.9 Text mining, méthodes d'analyse des sentiments/d'opinions et outils de visualisation des données

2.9.1 Text mining

L'exploration de texte ou text mining est un processus automatisé de la détection de nouvelles connaissances non découvertes dans des ressources énormes de textes non structurées. La recherche d'information (IR) utilise des requêtes de recherche ou par le biais des mots clés. Tandis que le text mining combine des techniques de la recherche d'information (RI), l'apprentissage automatique, les statistiques et le Traitement Automatique de la Langue Naturelle (TALN).

La Recherche d'Information (en anglais Information Retrieval) est définie par l'auteur [26] comme une activité d'un individu qui cherche à localiser et traiter une ou plusieurs informations au sein d'un document complexe, en vue de répondre ou résoudre un tel problème. Elle est définie aussi comme un ensemble d'action, méthodes et procédures de récupération d'informations sur un sujet donné ou références à des documents pertinents à partir de données stockées.

L'auteur [27] a subdivisé les principales méthodes utilisées en web pour la recherche d'information en quatre catégories :

- Recherche par mots clé et critères booléens

Tableau 7 Critères booléens dans la recherche d'information utilisant le moteur de recherche Google

Fonction de critère	Critère utilisé sur Google
Synonyme	~
À l'exception du mot (Sauf)	- Mot_clé
Ou	Mot_clé1 ou mot_clé2
Et	Mot_clé1 et mot_clé2 Mot_clé1 + mot_clé2 Mot_clé1 and mot_clé2
Définition	Define : mot_clé Définir : mot_clé
Dans le titre	Intitle :mot_clé

- Recherche assistée : Le moteur de recherche produit des suggestions ou recommandations basées sur la requête initiale de l'utilisateur.
- Recherche par classification : L'espace d'information est divisé en une hiérarchie de catégories, où l'utilisateur navigue du générique vers le spécifique.
- Requête par l'exemple : L'utilisateur sélectionne un élément intéressant d'un hypertexte, qui est ensuite utilisé comme base d'une nouvelle requête.

L'auteur [28] a ajouté une cinquième catégorie aux celles présentées ci-dessus de la recherche d'information, qui est la recommandation : en effet, cette dernière catégorie propose de l'information potentiellement intéressante en fonction de l'utilisateur et/ou du contexte de recherche.

Traitement Automatique de la Langue Naturelle (ou TALN) [29] est un sujet multidisciplinaire regroupant plusieurs axes de recherche tels que le linguistique, l'informatique et l'intelligence artificielle. Il est utilisé dans plusieurs champs et applications à savoir :

- Syntaxique (lemmatisation, racinisation, morphologie...)

Stemming ou racinisation : c'est le traitement qui permet d'aboutir à la racine d'un mot, elle est identifiée par la suppression du préfixe et du suffixe d'un mot. Tandis que lemmatisation est le fait d'obtenir la forme canonique d'un mot selon les deux cas suivants :

- ❖ Le mot se transforme en infinitif dans le cas d'un verbe
- ❖ Le mot se transforme en masculin singulier dans le cas d'un adjectif ou d'un adverbe.

Les exemples suivants illustrent la différence entre la racinisation et la lemmatisation :

- Racinisation : transformer les mots optimiser, optimal, optimisation en optim
 - Lemmatisation : transformer les mots définir, définition, défini, définissons, définie dans sa forme infinitive (définir).
- Sémantique (traduction automatique, génération automatique des textes, résumé automatique, correction orthographique...)
 - Traitement de signal (reconnaissance automatique de la parole, traitement de la parole...)
 - Extraction d'information (fouille de texte, recherche d'information, classification et catégorisation automatique des documents, analyse des sentiments...)

La figure suivante résume le processus d'exploration de texte (Text mining) :

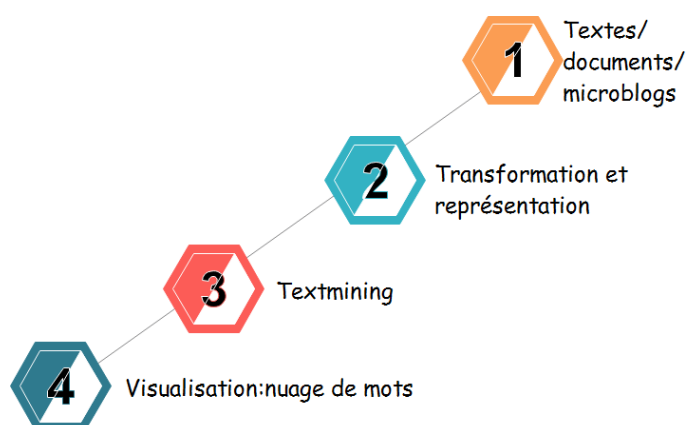


Figure 41 Processus de textmining et de visualisation des textes en nuage de mots

Dans ce qui suit, on va présenter les étapes de réalisation du processus schématisé dans la figure ci-dessus :

❖ Collecte de texte

Dans cette phase, on collecte les textes à partir de n'importe quelles sources de données (fichiers plats, fichier Excel, réseaux sociaux...)

❖ Prétraitement

Il est nécessaire de faire un prétraitement en vue de ne pas représenter tous les mots d'un corpus de texte ou de documents. Dans cette phase, on va supprimer les caractères spéciaux (&

] * etc...) et tous les mots vides (les articles, les pronoms, les prépositions...), des nombres, de la ponctuation, des espaces supplémentaires, des hashtags (s'il s'agit d'un corpus des tweets) ainsi que d'autres mots indésirables. Puis, on va appliquer le modèle d'implémentation très connu nommé algorithme de Porter, qui a été élaboré par l'auteur [30]. Cet algorithme a pour objectif d'effectuer un désuffixation.(Figure 42)

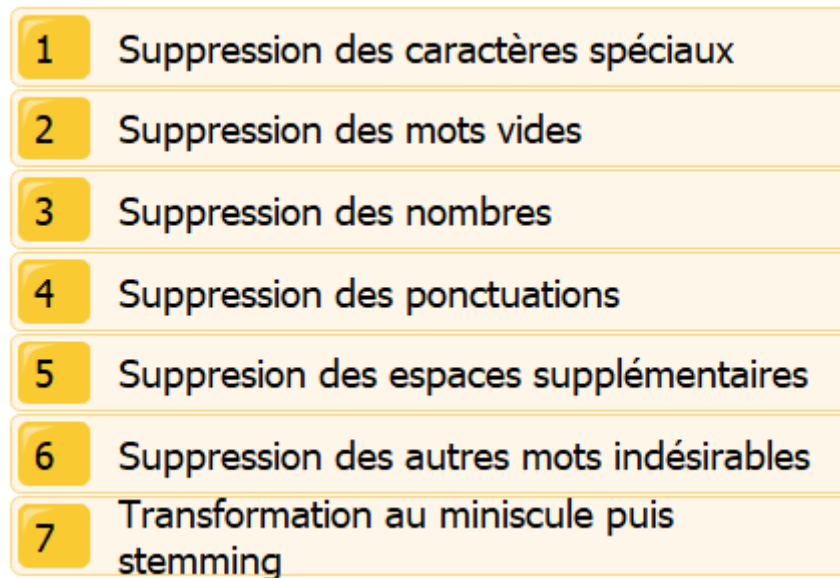


Figure 42 : Processus des transformations subites par le corpus de texte

❖ Création d'un corpus texte

Chaque colonne représente un document et chaque ligne représente une occurrence de terme dans tout le document (TF-IDF est lisible à l'annexe 2).

❖ Application des techniques d'exploration de texte (text mining)

On peut utiliser les méthodes exploratoires de regroupement, de classification et des règles d'associations (Kmean, CAH, KNN, SVM, ANN, ACP, ACM...) déjà présentées au début de ce chapitre.

❖ Visualisation

Dans cette phase finale, on peut adopter des outils de visualisation tels que le nuage de mots en vue d'extraire une information utile à partir d'une source d'informations assez volumineuses.

2.9.2 Analyse d'opinion ou des sentiments et outils de visualisation de texte

a) Analyse d'opinion ou de sentiments

Ces dernières années, l'information a connu une croissance quasi exponentielle en raison du développement des nouvelles technologies de l'information et de la communication et les réseaux sociaux (Twitter, Facebook..) deviennent non seulement comme un moyen de communication personnel, mais aussi comme un moyen de partager des opinions sur les produits et services ou même les événements politiques. Par conséquent, les entreprises ont été obligées de changer la manière d'analyser les perspectives et la satisfaction des clients en termes de leurs services et leurs produits puisque la collecte des opinions des clients peut être une tâche coûteuse et fastidieuse en utilisant des méthodes conventionnelles telles que les enquêtes traditionnelles.

L'analyse d'opinion (analyse d'opinion ou analyse de subjectivité) peut être définie comme un processus permettant de déterminer la polarité des opinions (positive, neutre ou négative) exprimé par les êtres humaines, elle peut être appliquée à toute forme textuelle d'opinions (blogs, textes de documents, revues, sondages et microblogs ...).

L'objectif de L'analyse des opinions des clients par les entreprises réside dans le fait de mieux comprendre leur valeur concurrentielle dans un marché en évolution et leurs avis sur les produits et services, ce qui leur permet également de prédire des futures stratégies marketing et politiques décisionnelles.

Elle est particulièrement utilisée en marketing pour analyser par exemple les commentaires des internautes ou les comparatifs et tests des blogueurs ou encore les réseaux sociaux : une grande part de la littérature sur le sujet concerne par exemple les tweets [31, 32,34 et 35]. Mais elle peut également être utilisée pour sonder l'opinion publique sur un sujet, pour chercher à caractériser les relations sociales dans les forums ou encore pour vérifier si Wikipédia est bien un média neutre.

Le tableau 8 ci-dessous présente quelques travaux traitant l'analyse des sentiments ou d'opinions :

Tableau 8 Quelques applications utilisant l'analyse des sentiments

Auteurs	Méthode
[31] et [32]	Extraire des opinions de Twitter sur les candidats aux élections présidentielles et prédire les résultats des élections
[33]	Fournir un modèle pour prédire les revenus attendus pour un film en analysant les données des médias sociaux
[34]	Analyse d'opinion sur un produit(téléphones Samsung Galaxy) en analysant les tweets publics sur Twitter
[35]	Utiliser les données de Twitter pour évaluer le sentiment des consommateurs de grandes marques telles que Nokia, Samsung, IBM et des compagnies aériennes telles que Egypt Air
[36]	Explorer les émotions dans les e-mails et les livres d'histoires.

L'analyse de sentiment peut être effectuée en utilisant une approche basée sur le lexique, une approche d'apprentissage automatique, et finalement une approche hybride [37]. La première approche est très populaire dans laquelle le lexique est composé d'un ensemble de mots d'opinion positifs et négatifs, utilisés pour noter les phrases d'opinion soit positive, négative ou neutre.

La figure suivante illustre l'approche basée sur le lexique d'opinion :

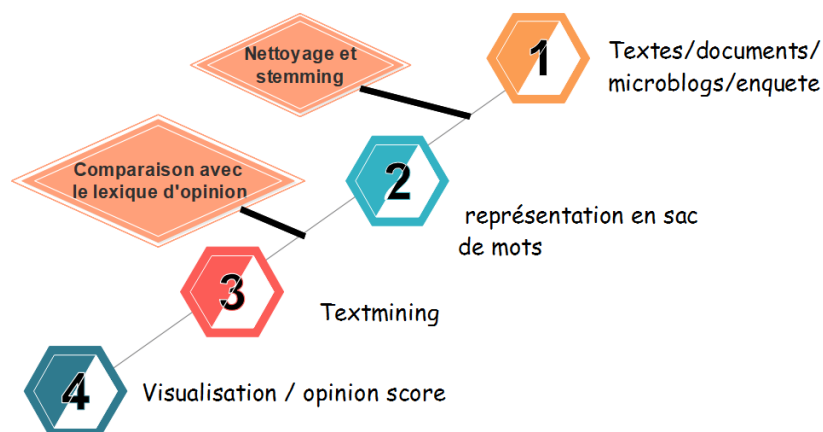


Figure 43 Processus d'analyse d'opinion

La deuxième approche utilise des techniques d'apprentissage automatique (KNN, ANN, Bayes naïf...) pour former un classificateur en utilisant un ensemble d'opinions pré-classifiées

comme une base d'apprentissage. Ensuite, utiliser le classificateur formé pour classer les nouvelles opinions comme positives, négatives ou neutres.

L'approche hybride c'est la combinaison des deux approches précédentes.

b) Outils de visualisation de texte : Nuage de mots

Le nuage de mots ou tag de mots fournissent un moyen simple et efficace pour communiquer visuellement les mots les plus fréquents des documents textes. Il est également appelé Word Cloud ou tag Cloud en anglais.

Aujourd'hui, le nuage de mots à plusieurs utilisations dans divers domaines. La figure suivante illustre quelques-unes de ses utilisations :



Figure 44 Quelques cas d'utilisation des nuages de mots

Les nuages de mots ont été utilisés pour la visualisation de texte dans de nombreuses études à des fins différentes. Par exemple, les auteurs [38] présentent des méthodes pour réduire et équilibrer l'espace blanc dans les nuages de mots en utilisant des dispositions rectangulaires pour améliorer l'affichage des nuages d'étiquettes.

Les auteurs [39] examinent divers problèmes techniques et d'évaluation de l'utilisabilité pour les quatre algorithmes basés sur les nuages d'étiquettes. Ils ont proposé une nouvelle famille d'algorithmes de disposition pour les nuages d'étiquettes. Ces algorithmes sont capables de produire des mises en page qui inscrivent des balises dans des polygones convexes arbitraires. D'autres algorithmes de mise en page ont été proposés dans le travail des auteurs [40] et [41]. Les auteurs [40] font une étude comparative entre wordle (cf. figure 45f) et manywordle (cf. figure 45g). Ils suggèrent que ManiWordle offre une plus grande satisfaction de l'utilisateur. Les auteurs [41] ont proposé un nouvel algorithme (Rwordle ou Rolled-out Wordles) et le

comparent aux algorithmes existants. Ils ont mentionné que cet algorithme préserve mieux la forme globale que les techniques connues.

D'autres conceptions avancées sont disponibles dans les générateurs de nuages de mots en ligne tels que Wordle, Tagul, Tagxedo, tocloud et makecloud. [42]

Les stratégies de mise en page vont de simples approches ligne par ligne (cf. figure. 45e) [43] à des mises en page de type treemap [44] et des placements forcés en combinaison avec des diagrammes de Venn (tagcluster (cf. figure.45a) [45]

Les auteurs [45] ont proposé une visualisation améliorée nommée TagClusters basée sur une analyse sémantique, les tags étant regroupés en différents groupes sémantiques. Leur distance visuelle dépend de la similarité sémantique entre les balises, et ainsi la visualisation offre une meilleure compréhension sémantique des balises générées en collaboration.

Les travaux [46] et [47] ont combiné des caractéristiques de projections multidimensionnelles (2D) et des nuages de mots dans un environnement visuel unique. Le premier travail fournit un mécanisme permettant la distribution sémantique des nuages de mots dans les polygones généraux tout en combinant cette sémantique avec l'information de pertinence et avec l'indexation des documents sous-jacents (cf. figure 45i). Tandis que le second travail proposait un nouvel algorithme pour créer des nuages de mots sains et préservant la sémantique en utilisant une technique de sculpture de couture adaptée (cf. figure 45h). Les auteurs [48] utilisent un mot topographique des paysages pour afficher des nuages de tags à grande échelle.

Les auteurs [49] utilisent des arborescences de préfixes pour regrouper différentes formes de mots (cf. figure. 45j), tandis que Word Cloud Explorer utilise un traitement Naturelle de la langue avancé pour lier des formes de mots et pour soutenir l'analyse visuelle de documents textuels via des nuages de mots interactifs. [50]

Cependant, tous ces travaux ne distinguent pas les différents documents texte, mais les mots sont traités comme s'ils provenaient tous de la même source de texte.

Les auteurs [51] présentent ConcentriCloud, une disposition de nuages de mots en couches qui fusionne les mots de plusieurs documents de texte en une seule visualisation. Les mots pondérés sont disposés selon une disposition concentrique, ceux qui représentent les documents individuels sur le cercle extérieur et ceux qui sont fusionnés sur les cercles intérieurs. (cf figure.

45b), l'auteur présente la visualisation ConcentriCloud des sept romans "Harry Potter" (HP1 à HP7).

Les auteurs [52] présentent une approche nommée word storm pour la comparaison visuelle des documents. Les résultats de cette approche confirment que, contrairement aux nuages de mots standards, une tempête de mots coordonnée permet une meilleure comparaison visuelle des documents. En effet, pour améliorer cette comparabilité, les mots apparaissant dans plusieurs documents sont placés au même endroit à travers les nuages, avec la même couleur et la même orientation. Les auteurs [53] utilisent une visualisation sunburst pour indiquer une hiérarchie de concepts extraits de documents texte (cf. Figure 45d).

Les auteurs [54] présentent l'approche RadCloud comme une technique de visualisation de texte basée sur des nuages de mots multiples fusionnés en une seule vue et les visualise dans un nuage de mots circulaires (cf. figure 45c). Les mots sont disposés radialement dans une disposition sans chevauchement. Les sources de texte sont indiquées par l'arrangement de mots spatiaux et les diagrammes à barres empilés...

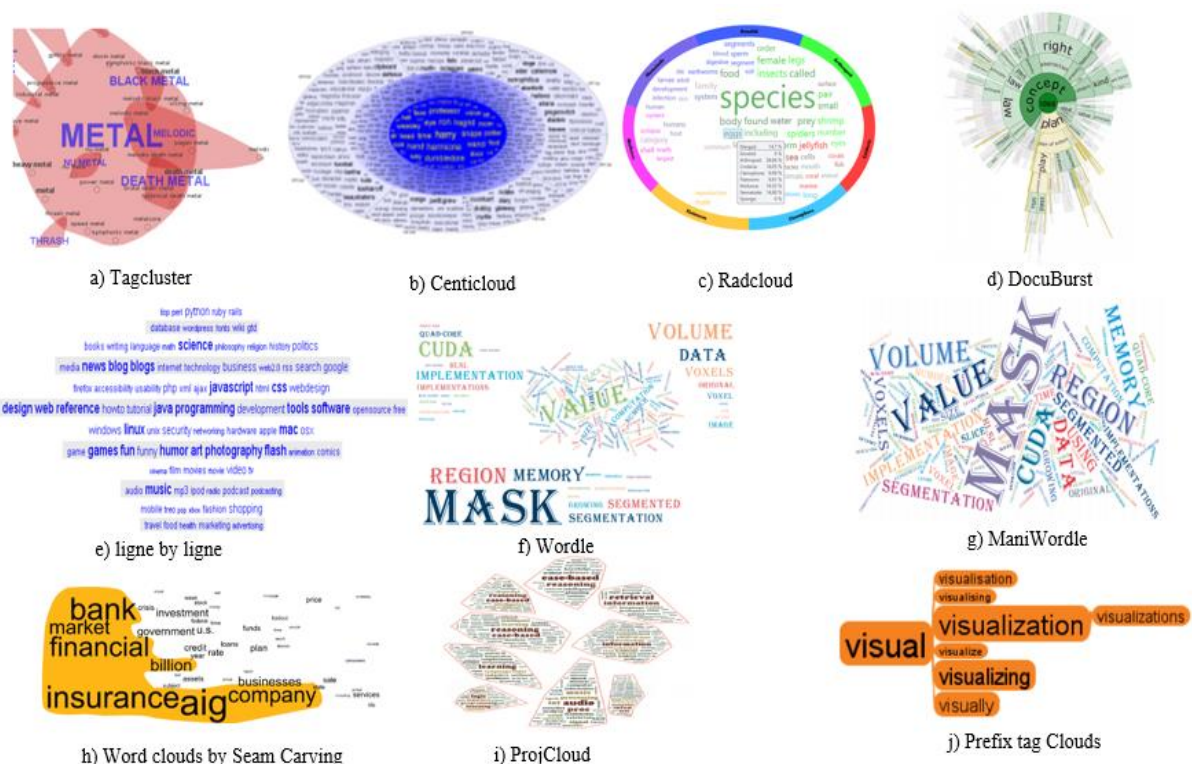


Figure 45 Différentes approches de visualisation du nuage de mots

2.10 Limites du système d'information relationnel

Les systèmes d'information décrites ci-dessus ont des limites pour les raisons suivantes :

- L'évolution d'espace de stockage, de production de l'information ;
- Le besoin d'y accéder de manière plus rapide et pertinente.

L'existence de ces limites mène à l'apparition des nouvelles modèles d'analyse de données autres que le modèle relationnel parmi lesquels le modèle multidimensionnel et le noSQL.

2.11 NoSQL et Big Data

Le nosql (not only sql) désigne une catégorie de base de données apparue en 2009 [55] qui se différencie du modèle relationnel que l'on trouve dans des bases de données connues comme MySQL, PostgreSQL, Oracle, Access... Elle est apparue pour remédier les inconvénients des Base de Données Relationnelles présentées dans la section précédente. Dans le marché des SGBD NOSQL [56] , on peut trouver les produits suivants :

- Cassandra, Hbase, CouchDB pour la fondation Apache ;
- MongoDB pour Mongo Inc ;
- Neo4j pour Neo Technology Inc.
- OrientDB pour OrientDB Ltd ;
- Oracle nosql pour Oracle corporation.

Afin de traiter une quantité énorme d'informations dans des délais plus courte, il est nécessaire d'améliorer nos plates-formes ainsi que les techniques d'analyse de données.

Nous avons adopté le Big Data pour les raisons suivantes :

- Augmentation rapide du nombre d'informations (des documents, des messages électroniques, des tweets, chat...) ;
- Nombre de transaction devient très important ;
- Des bases de données réparties sur plusieurs continents ;
- Le souhait de développer le chiffre d'affaires : mieux analyser les incidents passés pour anticiper l'avenir ;

- Prévention des risques ;
- Amélioration l'engagement des consommateurs ;
- À mieux cibler de nouveaux clients et à re-cibler ceux existants.

Dans la littérature, on a plusieurs définition de la nouvelle notion de Big Data. L'auteur [57] décrit les données massives par les 3 premières lettres V (3Vs) comme elles ont présenté dans la figure 46 :

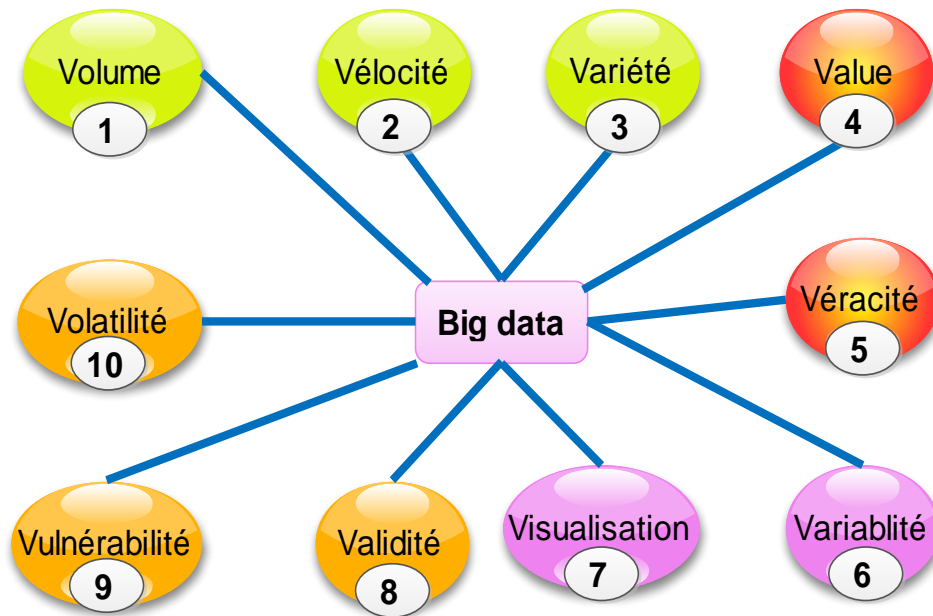


Figure 46 Les 10 Vs de Big Data

La figure 46 illustre les 10Vs de Big Data qui sont :

- Volume : désigne génération des données assez volumineuses (partage de photos des vidéos sur YouTube, Facebook, Twitter...)
- Vélocité : est la vitesse à laquelle les données sont générées, produites, créées ou rafraîchies mesuré en quantité d'informations par durée.
- Variété : Récemment, l'augmentation remarquable des sources de génération de données (réseaux sociaux, différentes applications, ERP...) provoque la diversité des données telles que le texte, la vidéo, l'audio et les images de grande taille.

L'auteur [58] a ajouté deux autres caractéristiques (Value et véracité) à la définition des données massives, d'où l'appellation 5Vs.

- **Value (Valeur) :** La valeur des données représente en effet des données volumineuses. Avoir des quantités continues de données n'est pas utile jusqu'à ce qu'il puisse être transformé en valeur. Les avantages et les coûts (Valeur) de l'analyse et de la collecte des données volumineuses sont des éléments plus importants lors de l'analyse des données volumineuses.
- **Véracité :** concerne davantage la provenance ou la fiabilité de la source de données, son contexte et la signification des données pour l'analyse basée sur celle-ci.

L'auteur [59] à leur tour, il a ajouté deux autres propriétés à la définition des données massives, il s'agit de la Variabilité et de la Visualisation.

- **Variabilité :** La variabilité fait référence à la cohérence des données et à la valeur des données.
- **Visualisation :** La visualisation est utilisée pour symboliser les grandes données dans une vue complète et déterminer les valeurs cachées. La visualisation est une clé essentielle pour faire du Big Data une partie intégrante de la prise de décision. Les outils de visualisation de Big Data actuels sont confrontés à des défis techniques (limites de la technologie en mémoire et de la faible évolutivité, de la fonctionnalité et du temps de réponse...)

Nous avons 3 autres caractéristiques attribuées à la notion de Big Data. Elle s'agit de :

- **Validité :** Il est essentiel de s'assurer que les données sont précises pour une utilisation future. Afin de prendre les bonnes décisions à l'avenir, les organisations devraient valider les données de façon notable.
- **Vulnérabilité :** Les données massives ont pris en compte de nouvelles préoccupations en matière de sécurité.
- **Volatilité :** en raison de la vitesse et du volume des données volumineuses, les entreprises doivent être soigneusement prise en compte la volatilité des données. En effet, Ces entreprises ont tendance de garder la disponibilité des données à une date indéterminée en assurant une récupération rapide des informations chaque fois que cela est nécessaire à moindre coût.

2.12 Conclusion

Dans ce deuxième chapitre, nous avons tout d'abord abordé l'évolution des systèmes d'information de la base données relationnelle au Big Data, nous avons cité les méthodes d'analyse des données existe dans la littérature (l'analyse descriptive, l'analyse exploratoire, les outils d'analyse d'opinions et les outils de visualisation de données) et finalement nous avons présenté les caractéristiques des données massives. Le chapitre suivant abordera l'état de l'art de l'expérimentation en commençant par une étude de Benchmarking des outillages servant à la gestion des documents et de contenu.

Chapitre 3. Benchmarking des outillages

Chapitre 3. Benchmarking des outillages

3.1 Introduction

Dans ce chapitre, on entamera trois études comparatives des outillages : d'une part l'étude comparative des Systèmes intégrés de gestion de bibliothèque (SIGB), une étude comparative des ERP disponibles sur le marché et une autre étude concernant l'évaluation des outils d'analyse de données massives , et d'autre part on présente les critères de choix pour chaque étude ainsi que les solutions retenues pour accomplir notre travail.

3.2 Benchmarking de l'outillage SIGB

Le but de cette section est d'évaluer les logiciels de bibliothèque numérique afin de justifier notre choix pour déterminer quels logiciels offrent plus en termes de services et de facilité d'utilisation. Nous avons utilisé une étude comparative avec les caractéristiques de logiciels en fonction de leur finalisation et de leur maturité. Dans la suite de cette étude comparative le signe – signifie que la propriété ou la fonctionnalité est manquante ou non reconnue.

La figure 47 illustre le processus de Benchmarking des outillages.

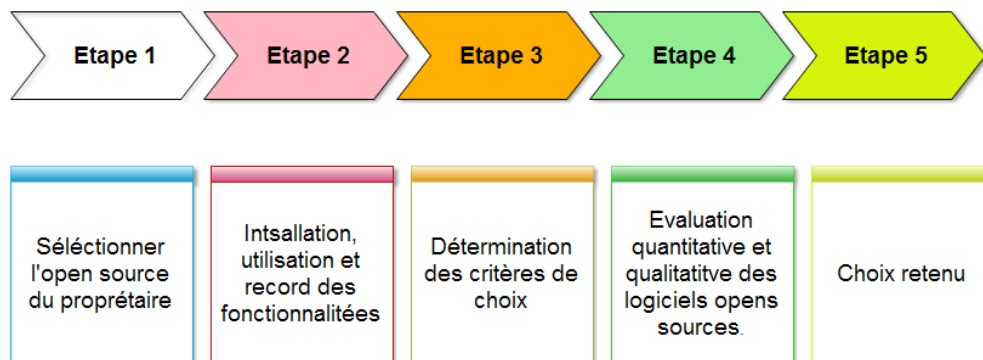


Figure 47 Processus de Becnhmarking des outillages

3.2.1 Solutions gratuites

Les tableaux 9 à 13 présentent la liste des propriétés des SIGB gratuites :

Tableau 9 Liste des SIGB gratuites

Logiciel	Langage de programmation	Base de données	Licence	Web	Multiplateforme	SIGB : oui CMS : Non
Biblioteq	C++	Postgresql,sqlite	BSD	-	Oui	Oui
Greenstone	C++,Perl,Java	-	GPL	Oui	Oui	Oui
Emilda	C++,PHP, Perl	Mysql	GPL	Oui	Oui	Oui
Fedora commons	Java	-	Gpl	-	Oui	Oui
Dspace	Java	Postgresql,Oracle	BSD	Oui	Oui	Oui
Kuali OLE	Java	Mysql,oracle	GPL	Oui	-	Oui
Newgenlib	Java,JDOM	Postgresql	GPL	Oui	Oui	Oui
Alfresco community	Java, jsp et javascript	Mysql,Postgresql	Gpl	Oui	Oui	Non
Cds/isis	Pascal	-	Unesco	-	Oui	Oui
Eprints	Perl	Mysql	GPL	Oui	Oui	Oui
Koha	Perl	Mysql	GPL	Oui	Non	Oui
Mylibrary	Perl	Mysql,Postgresql	Gpl	Oui	Linux	Oui
Evergreen	Perl, C, Xulrunner	Postgresql	GPL	Oui	Non	Oui
Opals	Perl, java	Mysql	Gpl	Oui	Linux	Oui
ABCD	PHP	MySQL	GPL	Oui	Oui	Oui
Bibliotecaphp	PHP	-	GPL	Oui	Oui	Oui
Drupal opac	PHP	MySQL	GPL	Oui	Oui	Oui
Drupal sopac	PHP	MySQL	GPL	Oui	Oui	Oui
Extensible Catalog	PHP	MySQL	GPL	Oui	-	Oui
Ganesha	PHP	MySQL	-	Oui	Oui	Oui
Keystone	PHP	MySQL	GPL	Oui	Non	Oui

Logiciel	Langage de programmation	Base de données	Licence	Web	Multiplateforme	SIGB : oui CMS : Non
Openbiblio	PHP	MySQL	GPL	Oui	Oui	Oui
Otomigenx	PHP	MySQL	GPL	Oui	Oui	Oui
Phpmybibli : PMB	PHP	MySQL	GPL	Oui	Oui	Oui
Potnia	PHP	-	MPL1.1	Oui	-	Oui
Refbase	PHP	MySQL	GPL	Oui	Oui	Oui
Tematres	PHP	-	GPL	Oui	-	Non
Vufind	PHP	MySQL	GPL	Oui	-	Oui
Phpmylibrary	PHP,pl/sql,Python	MySQL	GPL	Oui	Linux	Oui
Invenio	Python	MySQL	GPL	Oui	-	Oui
Mycore	Xsl,Java	-	GPL	Oui	-	Oui

Dans ce qui suit, nous présentons les logiciels Open Source, multiplateforme et web based utilisant le langage de programmation PHP, qui participent à cette étude comparative, à savoir ABCD, OpenBiblio, PhpMyBibli et Potnia.

Le tableau 10 suivant présente la liste des propriétés des SIGB gratuites :

Tableau 10 Logiciels open sources de la gestion de bibliothèque

Logiciel et site web	Développeur	Licence	Dernière version
ABCD (https://sites.google.com/site/abcdlibraryautomationsoftware/downloads)	BIREME (WHO, Brazil)	GNU GPL	2.0 en 22 Septembre 2011
OpenBiblio (http://obiblio.sourceforge.net)	OpenBiblio Development Team	GNu GPL	0.7.2 en Aout 13, 2014
PhpMyBibli (http://www.sigb.net)	PMB Services	GNu GPL & CECILL	4.2.7 en juillet, 2016
Potnia (https://sourceforge.net/projects/potnia/)	Directorate General of Universities of the Spanish Ministry of Education, Culture and Sport	MPL1.1	1.0. 12, en Octobre 2003
BibliotecaPHP (http://bibliotecaphp.sourceforge.net/)	J.E. Amaro	Gnu GPL	1.1 en 4 Avril ,2010

Tableau 11 Langues supportées, documentation et taille du logiciel

Logiciel	Langages	Documentation	Taille
ABCD	Multilingue	Disponible en anglais, français et espagnol (http://wiki.bireme.org/en/index.php/ABCD_resources)	48.6Mo
OpenBiblio	Anglais et traduit en espagnol	Disponible en anglais(http://obiblio.sourceforge.net/)	759.1 KB
PhpMybibli (PMB)	Multilingue	Multilingues (http://www.pmbservices.fr/index.php?cms_build_activate=1&lvl=cmspage&pageid=20) in French PMB release 3.4	55.2MB
Potnia	Espagnol	Disponible en anglais https://sourceforge.net/projects/potnia/files/potnia/Potnia%201.0/	162.5 Kb
BibliotecaPHP	Espagnol	Multilingues (https://sourceforge.net/projects/bibliotecaphp/files/)	176.1 Kb

Tableau 12 Mode de développement des SIGB

Logiciel	Première Version	Dernière version	Plateforme	Base de données	Interface utilisateur	Langage de Programmation
ABCD	2011	2011	Multiplateforme	MySQL	Web-based	PHP
OpenBiblio	2002	2014	Multiplateforme	MySQL	Web-based	PHP
PhpMybibli	2003	2014	Multiplateforme	MySQL	Web-based	PHP
Potnia	2003	2013	Windows	MySQL	Web-based	PHP
BibliotecaPHP	2010	2011	Multiplateforme	-	Web-based	PHP

- ABCD:

ABCD (Figure 48) est l'acronyme d'une suite logicielle pour l'automatisation des bibliothèques et des centres de documentation. Le nom lui-même exprime déjà l'ambition de la suite logicielle de fournir non seulement des fonctions d'automatisation pour les bibliothèques traditionnelles, mais aussi d'autres fournisseurs d'informations tels que les centres de documentation. Il est développé par BIREME (OMS, Brésil) en collaboration avec le Conseil interuniversitaire flamand, Belgique, et en utilisant la technologie de base de données ISIS de l'UNESCO. Elle vise à fournir un outil de gestion de bibliothèque intégré, couvrant toutes les fonctions majeures d'une bibliothèque, comme les acquisitions, la gestion de bases de données bibliographiques, la gestion des utilisateurs, les transactions,

le contrôle en série, la recherche en ligne des bases de données bibliographiques locales et externes, portail de la bibliothèque. Les principales caractéristiques d'ABCD sont la couverture des principales fonctions de la bibliothèque, sa centralité web et son développement et sa maintenance dans le cadre de la méthodologie des logiciels libres et Open Source.

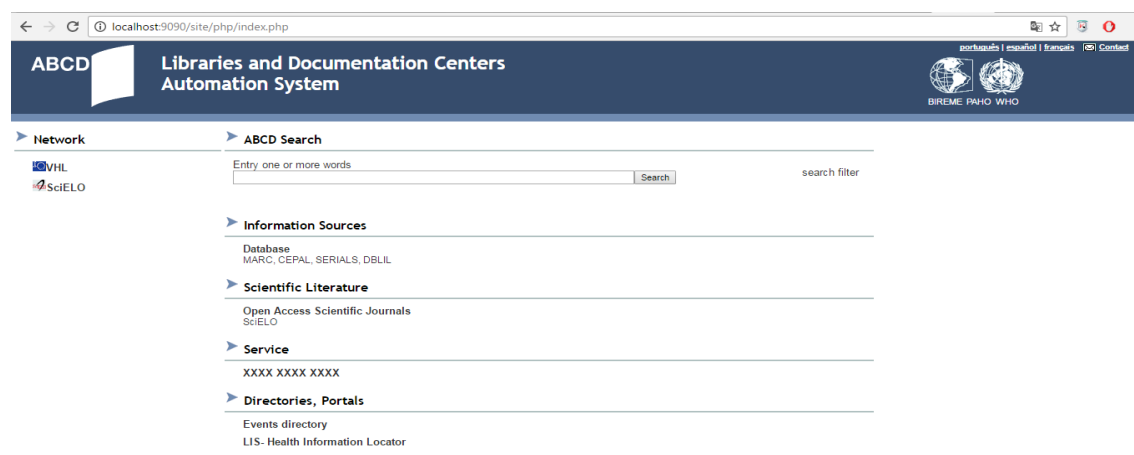


Figure 48. Interface utilisateur OPAC pour ABCD

- OpenBiblio:

OpenBiblio (Figure 49) est un système de gestion de bibliothèque intégré, il a été créé en 2002 par Dave Stevens sous GNU GPL. Il a été conçu pour les petites bibliothèques, surtout les bibliothèques scolaires. Ce système a été traduit en espagnol appelé EspaBiblio.

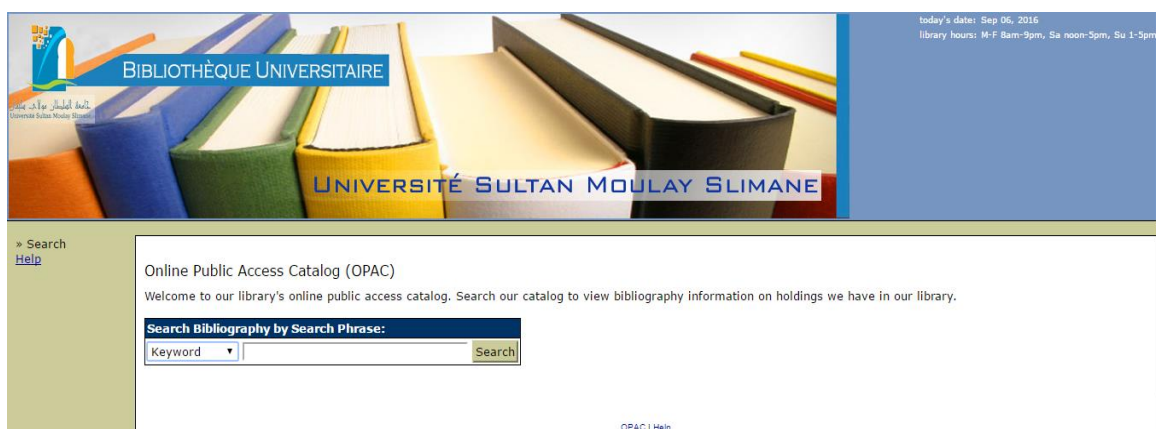


Figure 49. Interface utilisateur OPAC pour OpenBiblio

- PhpMyBibli:

PhpMyBibli ou PMB (Figure 50) est un système de bibliothèque intégré open source. Il est développé et maintenu par la société française PMB Services en 2002 sous licence CeCILL. PMB est conçu pour les bibliothèques de taille moyenne et grande, et peut également gérer des

réseaux de bibliothèques intégrés dans un catalogue collectif. Le format de catalogage utilisé est UNIMARC mais il intègre des systèmes de conversion et d'importation de formats (USMARC et XML, entre autres).

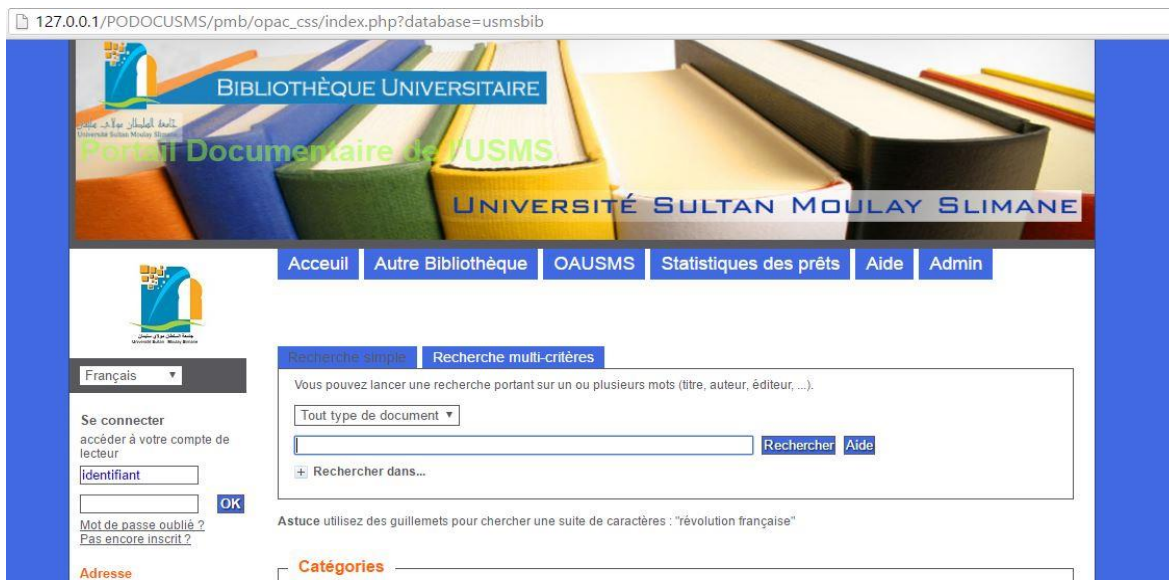


Figure 50. Interface utilisateur OPAC pour PMB

- Potnia:

Potnia (Figure 51) crée des répertoires thématiques, pour la gestion des ressources scientifiques, des revues, des publications. Il a été créé en 2003 par la Direction Générale des Universités du Ministère de l'Éducation, de la Culture et du Sport d'Espagne. Il a une interface en langue espagnole.



Figure 51. Interface utilisateur OPAC pour Potnia

- Bibliotecaphp:

Bibliotecaphp (Figure 52) est un site web open source. Programme de gestion d'une petite bibliothèque écrite en PHP sans base de données. Ce programme a été développé par J.E. Amaro dans un projet d'innovation pédagogique de l'Université de Grenade. [60]

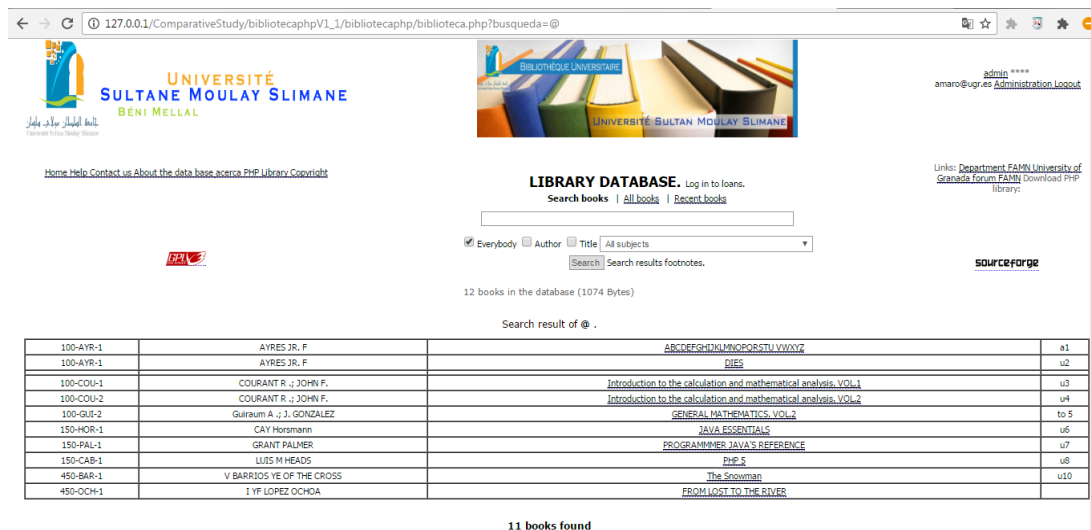


Figure 52. Interface utilisateur OPAC pour Bibliotecaphp

3.2.2 Solutions propriétaires

Le tableau 13 présente la liste des SIGB propriétaires :

Tableau 13 Liste des logiciels de SIGB propriétaires

Logiciel	Langage de programmation	Base de données	Licence	Web	Multiplateforme	SIGB : oui CMS : Non
Worldcat	-	-	Oclc	Oui	Oui	Non
E-library	C++, Javascript	-	GPL	-	Non	Oui
Insignia	Html5	-	-	Oui	Oui	Oui
Libsys	Java	Mysql	-	Oui	-	Oui
Symphony	Java, jsp	Oracle	-	Oui	Oui	Oui
Voyager	Perl script	-	GPL	Oui	Oui	Oui

3.2.3 Critères de choix et résultat

La plupart des bibliothèques des universités Marocaines ne peuvent pas acheter de logiciels commerciaux en raison de contraintes budgétaires sévères tels que le prix élevé du logiciel de bibliothèque. Par conséquent, elles expriment leurs besoins d'utiliser un logiciel Open source.

La numérisation des bibliothèques est importante car elle peut prendre en charge les différents besoins à faible coût et le code source est disponible pour des modifications et des améliorations.

Les auteurs [61] présentent les avantages de l'open source comme suit :

- L'utilisation sans restriction ;
- Le coût gratuit ;
- L'implication de la communauté dans le développement et la maintenance de logiciels ;
- La compétence par rapport à d'autres logiciels commerciaux ;
- Les problèmes de copyright, etc.

Pour un bon choix du SIGB, nous avons suivi la procédure suivante :

❖ *Étape 1 : Sélectionner des logiciels open source par rapport aux propriétaires*

- Le logiciel doit être téléchargeable gratuitement et disponible pour l'installation avec un nombre de bases installées raisonnables.
- Open source : avec son code source mis à disposition avec une licence dans laquelle le détenteur des droits d'auteur accorde le droit d'étudier, de modifier et de distribuer le logiciel.
- Web based: applications qui communiquent sur Internet avec l'utilisateur via http.
- Multiplateforme : le logiciel peut fonctionner sous Linux, Windows et Mac OS.
- Licence GPL : le logiciel est sous licence GPL.
- Le logiciel doit être un SIGB et non seulement un Système de Gestion de Contenu (Contenant Management System CMS en anglais).

❖ *Étape 2 : Évaluation des logiciels open sources selon des fonctionnalités*

Pour comparer ces logiciels, nous adoptons le système de notation défini par Tristan Muller [62] présenté dans le tableau 14 ci-dessous avec un grade d'échelle allant de 0 à 7, où 7 est la notation la plus élevée.

Tableau 14 Schéma de système de notation défini par Tirstan Muller

Spécification	Échelle	Performance
Non disponible	0	N'est pas évalué
Disponible	1	La mise en œuvre est médiocre - indésirable
	2	À juste acceptable
	3	Amélioration adéquate mais peu significative
	4	Bien
	5	Très bien
	6	Mise en œuvre excellente - peu d'amélioration
	7	Implémentation remarquable - pas de défaut observé

Le tableau 15 présente les principales caractéristiques des cinq logiciels open sources qui participent dans l'étude comparative (1 la fonction existe ; 0 est manquante).

Tableau 15 Principales caractéristiques des cinq logiciels open sources

Fonctionnalité	ABCD	OpenBiblio	PhpMybibli	Potnia	BibliotecaPHP
Administration	1	1	1	1	1
Circulation	1	1	1	0	0
Catalogage	1	1	1	1	1
OPAC	1	1	1	1	1
Rapports	1	1	1	0	0
Acquisition	1	0	1	0	0
Gestion des publications en série	1	0	1	0	0
Total	7	5	7	3	3
Pourcentage %	100	71.43	100	42.86	42.86

Les valeurs calculées présentées dans le tableau ci-dessus (Tableau 15) prennent deux valeurs 1 (fonctionnalité existe) ou 0 (Fonctionnalité est manquante). Tandis que celles présentées dans le tableau ci-dessous (Tableau 16) sont à la base de critère de Tirstan Muller allant de 0 à 7 (Tableau 14).

Tableau 16 Évaluation des cinq logiciels open sources de la gestion de bibliothèque

Caractéristiques	ABCD	OpenBiblio	PhpMybibli	Potnia	BibliotecaPHP
Administration	4	4	6	2	2
Recherche Avancée	4	0	6	4	2
Contrôle d'autorité	4	4	6	2	4
Recherche de base	6	3	6	4	4
Catalogage	5	4	6	1	2
Circulation	5	3	6	0	0
Gestion des publications en série	4	0	4	0	0
Documentation	2	2	2	1	2
Documents acquisitions	4	0	5	0	0
DSI	4	3	5	0	0
Exporter / importer des données	4	3	5	0	0
prêts inter Bibliothèque	0	0	0	0	0
Normes de métadonnées prises en charge	5	5	5	0	0
OPAC	4	3	4	2	3
Renouvellement	4	4	5	0	0
Rapport et statistiques	4	4	4	0	0
Réservation	5	0	6	0	0
Recherche sémantique	0	0	0	0	0
Partager sur les sites de réseaux sociaux	0	0	0	0	0
Une caractéristique sociale comme aimer, commenter	0	0	0	0	0
Prise en charge multilingue	4	2	6	1	2
Assistance aux utilisateurs	4	4	5	1	4
Protocole Z39.50	4	0	6	0	0
Total	80	48	98	18	25
Pourcentage %	51.95	31.17	63.64	11.69	16.23

❖ *Étape 3 : Choix de la solution retenue*

Selon cette étude, PMB (PhpMybibli) était une solution open source plus appropriée pour le management des bibliothèques.

Le choix d'utiliser le SIGB PMB est due à l'utilisation facile du logiciel, les fonctionnalités qu'il offre et des raisons économiques.

3.3 Benchmarking de l'outillage des ERP

Un ERP (Enterprise Resource Planning), signifiant littéralement en anglais, planification des ressources de l'entreprise, et traduit en français par progiciel de gestion intégré (PGI) est un progiciels qui permet de gérer l'ensemble des processus d'une entreprise en intégrant l'ensemble des fonctions de cette dernière comme la gestion comptable, commerciale, financière, vente...Il intègre tous les systèmes informatisés permettant de soutenir le bon fonctionnement de l'entreprise.

3.3.1 Solutions gratuites

Aujourd'hui, les ERP gratuit trouvent leur place dans le marché pour les raisons suivantes :

- L'évolution du nombre de programmeur ;
- La vulgarisation des ordinateurs et des appareils mobiles ;
- L'accès à internet ;
- Le faible coût.

Le tableau 17 présente les caractéristiques des différents ERP open sources :

Tableau 17 Quelques caractéristiques des ERP open sources

Logiciel open source	V1	VF	licence	Plateforme	Base de données	Langage de programmation	Site web
Adempiere	2006	2017	Oui	Multi plateforme	PostgreSQL, Oracle	Java	http://www.adempiere.net/
Compiere	2002	2014	GPL, Propriétaire	Unix, Windows	multiple base de données	Java	http://www.compiere.org/
Dolibarr	2002	2018	GPL	Multi plateforme	MySQL, PostgreSQL	PHP	http://www.dolibarr.fr/
Ekylibre	2005	2018	GNU/AGP Lv3	Multi plateforme	PostgreSQL/ PostGIS	Ruby, Rails framework	http://ekylibre.org/
ERP5	2002	2014	GPL	Multi plateforme	MySQL	Python	https://www.erp5.com/

Logiciel open source	V1	VF	licence	Plateforme	Base de données	Langage de programmation	Site web
ERPnext	2008	2018	GPL	Multi plateforme	MariaDB	Python, JS	https://erpnext.com/
idempiere	2012	2018	GPL	Multi plateforme	PostgreSQL, Oracle	Java	http://www.idempiere.org/
EasyERP	2014	2016	Lgplv3	Windows		JS	https://easyerp.com/
Odoo	2004	2017	GPL	Multi plateforme	PostgreSQL	Python	http://www.odoo.com/
OfBiz	2010	2018	Apache License 2.0	Multi plateforme	Derby, MySQL, PostgreSQL	Java	https://ofbiz.apache.org/
OpenBravo	2005	2011	MPL	Multi plateforme	PostgreSQL, Oracle	Java	http://www.openbravo.com/
Openflex	2002	2018	GPL	Multi plateforme	MySQL	PHP	https://openflex.cloud/fr/
WebERP	2003	2018	GPL	Multi plateforme	MySQL	PHP	http://www.weberp.org
Xtuple	2005	2018	GPL	Multi plateforme	PostgreSQL	C++	https://xtuple.com/
OrangeHRM	2005	2018	GPL	Multi plateforme	MYSQL	PHP	https://www.orangehrm.com/

Dans ce qui suit, on présente quelques interfaces graphiques de quelques ERP participantes dans cette étude comparative :

La figure 53 présente le tableau de bord dans l'ERP Odoo, tandis que les figures 54 et 55 présentent respectivement la page d'authentification et son module de Gestion Électronique de Document (GED) de l'ERP dolibarr.

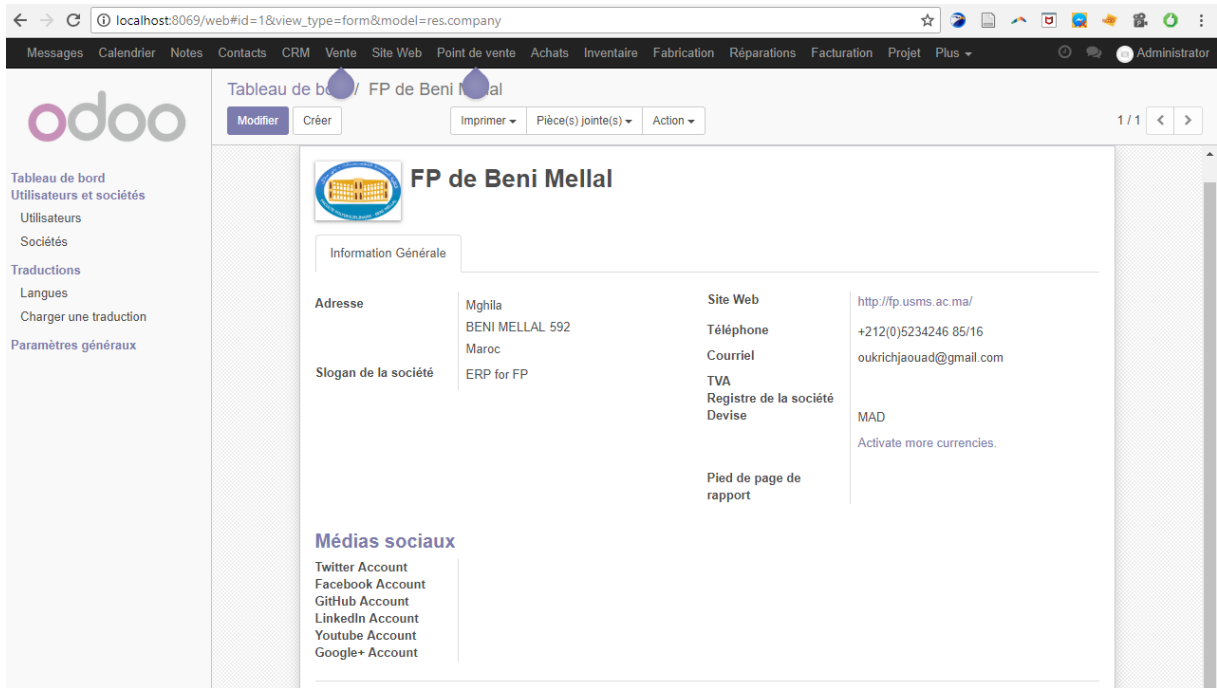


Figure 53 Interface Odoo 11

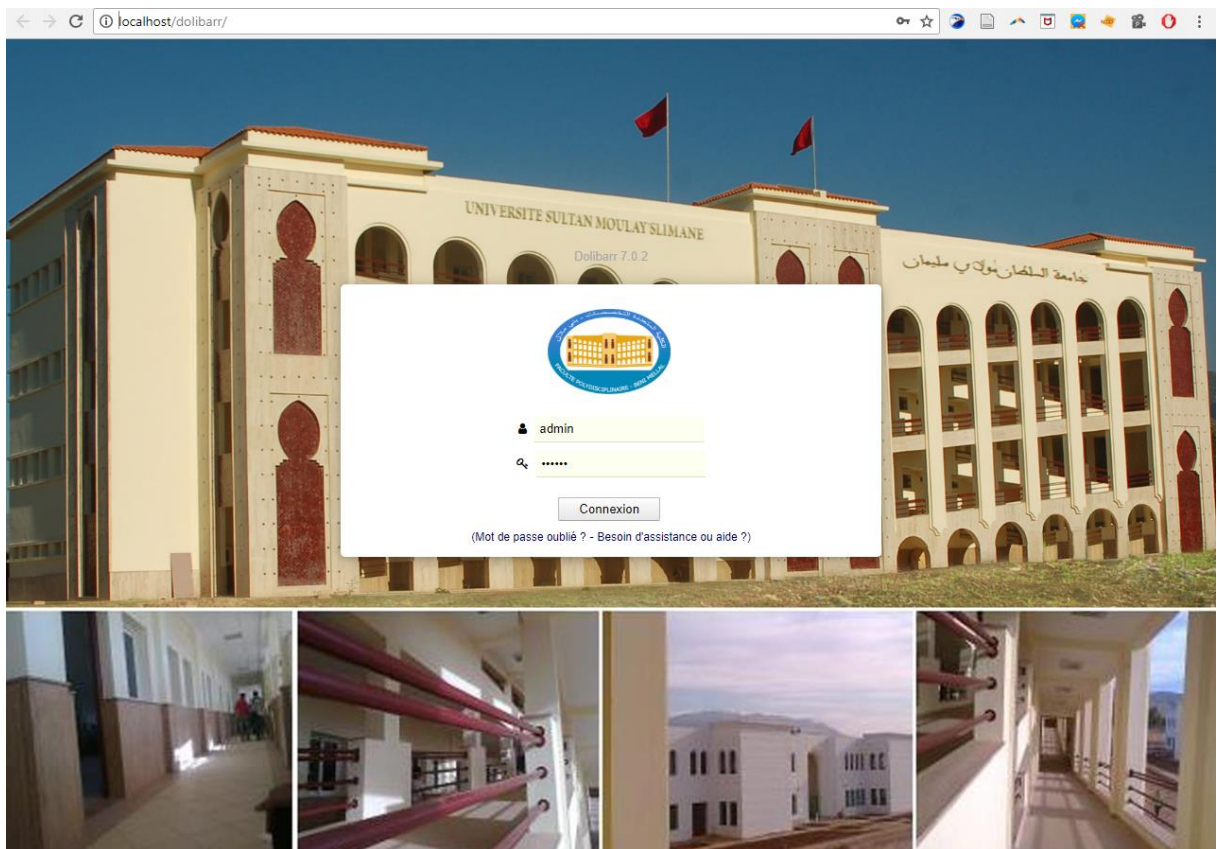


Figure 54 Interface d'authentification Dolibarr FP de Beni Mellal

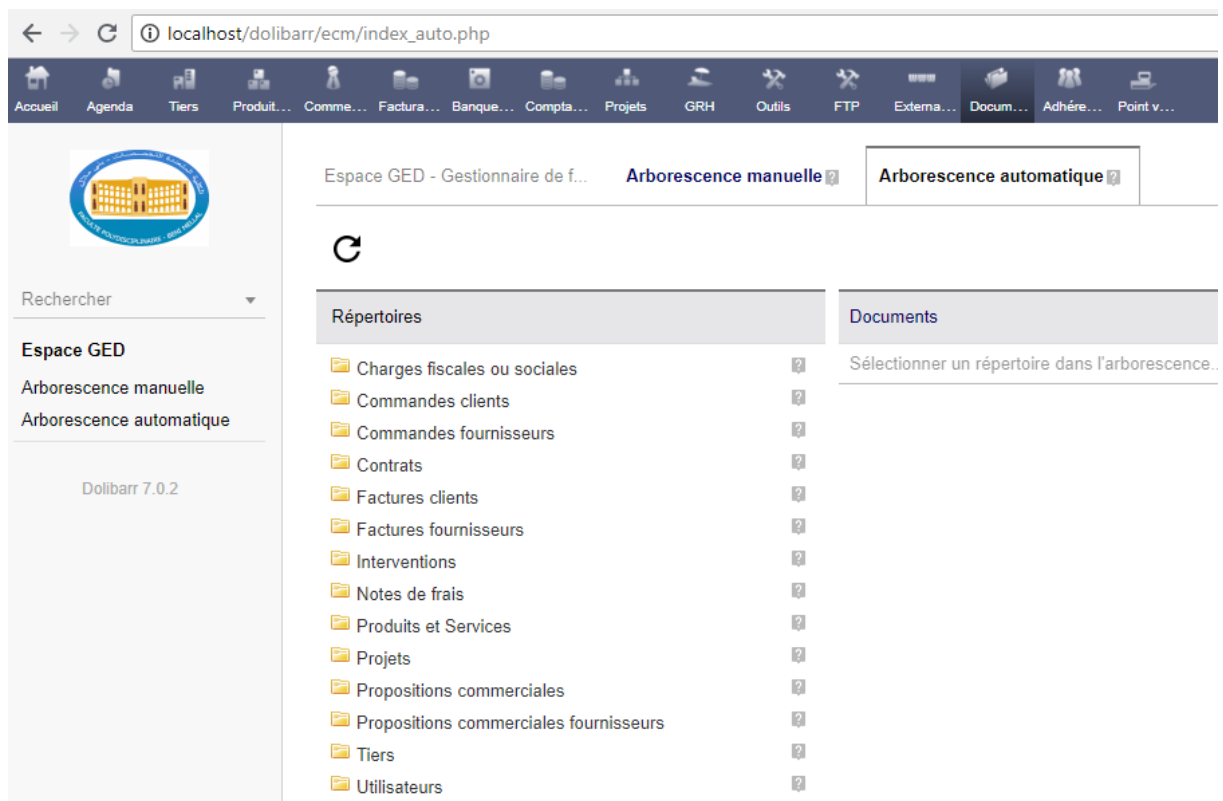


Figure 55 Gestion électronique des documents en Dolibarr

3.3.2 Solutions propriétaires

Parmi les solutions propriétaires existe au marché, on peut trouver Sage, oracle people soft... Les établissements Marocaines expriment leurs besoins d'adopter les solutions open sources pour des raisons économiques.

3.3.3 Critères de choix et résultat

Pour bien choisir notre ERP, il doit répondre aux caractéristiques suivantes :

- Il émane d'un concepteur unique ;
- En cas d'impact d'un module, l'information est mise à jour en temps réel dans l'ensemble des autres modules associés ;
- C'est un système qui garantit la piste d'audit donc il est facile de retrouver et d'analyser l'origine de chaque information ;
- Il peut couvrir l'ensemble du Système d'Information de l'entreprise ;

- Il garantit l'unicité des informations qu'il contient puisqu'il n'a qu'une seule base de données au sens logique ;
- Logiciel doit respecter la règle des 3S (Simple à développer, Simple à installer et Simple à utiliser) ;
- Logiciel doit être multiplateforme.

La solution retenue est Dolibarr pour les raisons suivantes :

- Très facile à utiliser même si vous n'êtes pas informaticien en utilisant l'auto installeur Doliwamp pour Windows ;
- Ergonomie agréable ;
- Multilingue ;
- ERP Gratuit ;
- Existence d'une communauté très active ;
- Langage de développement en PHP.

Parmi les limites de cet ERP :

- Il est encore faible d'importer une liste des clients ou des fournisseurs.

3.4 Benchmarking de l'outillage d'analyse de données massives (Big Data)

En analysant la littérature, Nous avons trouvé une liste des systèmes d'analyse de données massives parmi les plus fameuses on trouve le Framework Hadoop.

Plusieurs célèbres sociétés déclarent qu'ils utilisent Hadoop comme Facebook, Yahoo et Microsoft. [63]

Le logiciel Hadoop est une plate-forme logicielle open-source écrite en Java pour l'analyse de données massives. Il comprend un système de fichiers distribué HDFS, une plate-forme de traitement parallèle appelée Apache Hadoop MapReduce, et plusieurs composants prenant en charge l'importation des données, la coordination des flux de travail, la gestion des tâches, et le contrôle du cluster.

La figure 56 illustre l'architecture du Framework Hadoop :

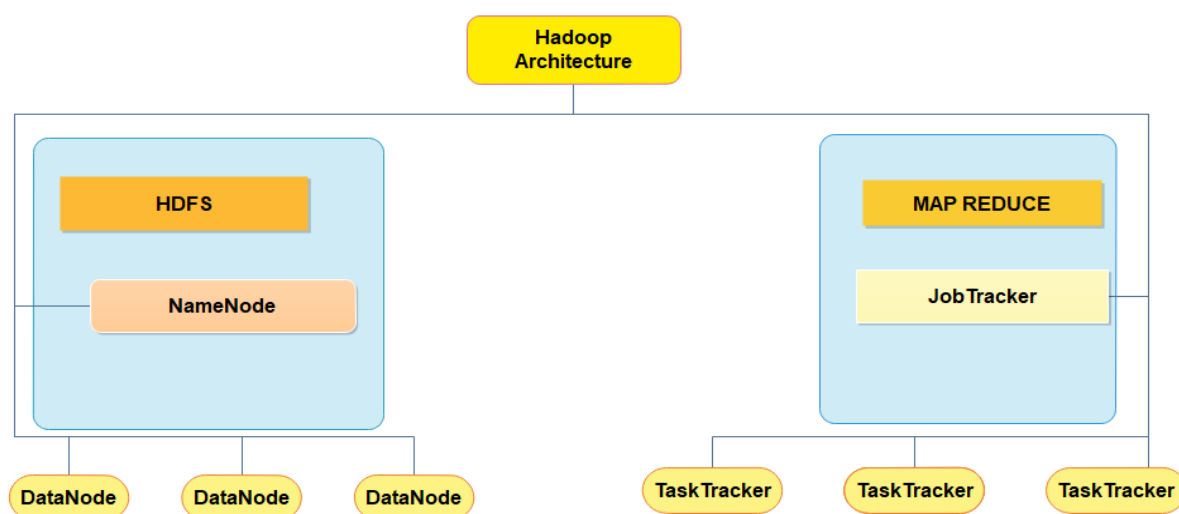


Figure 56 Architecture Hadoop

HDFS comporte deux types de composantes essentielles :

- NameNode (Nom des nœuds) : ce composant gère l'espace de nom, l'arborescence du système de fichier et les métadonnées.
- DataNode (Nœud de données): stock et restitue les blocs de données.

MapReduce est un Framework central du logiciel Apache Hadoop permettant un traitement de données massives en cluster. Il permet de décomposer le traitement de données en deux phases :

- Map : les données sont extraites par les mapper, les transformer puis les préparer pour les reducers.
- Reduce : permet de traiter et d'analyser les données extraites par les mapper.

Hadoop offre plusieurs avantages clés pour l'analyse de Big Data :

- Stockage de n'importe quel type de données ;
- Adaptabilité via des sociétés comme Facebook et Yahoo ;
- Offert de nouvelles perspectives en dépassant les approches traditionnelles de data mining ;

- Réduction des coûts : le stockage peut être ajouté progressivement en fonction des besoins, et le matériel peut être ajouté ou échangé au sein ou hors d'un cluster ;
- Une plus grande disponibilité en offrant une tolérance aux pannes grâce à la réplication des données et au basculement entre les nœuds de calcul ;
- Diminution du risque : la communauté Hadoop est active et diversifiée, avec des développeurs et des utilisateurs issus de nombreux secteurs à travers le monde.

Nous avons proposé d'adopter le système Hadoop HDFS comme un outil d'analyse de données massives parce que c'est un moyen plus économique pour traiter de grands ensembles de données NoSQL.

3.5 Conclusion

Les études comparatives des outillages nous a permis d'adopter les logiciels opens sources pour des raisons économiques et la possibilité de développement et de modification de ces logiciels.

Le chapitre suivant sera consacré à la description des approches proposées pour le management de l'université en utilisant les outillages retenus : PMB pour le management de la bibliothèque, ERP Dolibarr version 7.2 pour la gestion des transactions des processus de l'université, et hadoop HDFS pour analyser la totalité des données massives des différents systèmes intégrés de l'université et finalement le logiciel R pour l'analyse des opinions.

Chapitre 4. Contributions

Chapitre 4. Contributions

4.1 Introduction

Ce chapitre se consacrera à la description de nos contributions dans la gestion des documents et leurs contenus. En premier lieu, nous présenterons le management de la bibliothèque de l'établissement de l'USMS à travers l'instauration du système intègre PMB, ensuite, nous utiliserons le référentiel de connaissances par le biais des annotations des exemplaires et des notices bibliographiques. Puis, nous proposerons la contribution d'analyse d'opinion appliquée pour une enquête de satisfaction. Ensuite, nous présenterons une approche qui s'intéresse à l'intégration d'un progiciel de gestion intégré ERP dans le management de l'université. Et finalement nous donnerons une méthode d'analyse de données massives des différents systèmes intégrés au sein de l'Université.

4.2 Nouvelle stratégie de management documentaire de la bibliothèque local de l'USMS

4.2.1 Motivation et objectifs

Le Portail Documentaire numérique de l'Université Sultan Moulay Slimane a pour objectif de moderniser la gouvernance de l'établissement public de l'Université et à dépasser les limites des bibliothèques traditionnelles (l'absence de décision concernant des nouveaux achats, lentement de services, insatisfaction des lecteurs, l'ignorance et le manque des documents les plus utiles), pour répondre aux besoins des lecteurs, pour aider les dirigeants à prendre de meilleures décisions stratégiques et instantanés, et d'améliorer le rendement universitaire.

L'informatisation de la bibliothèque de l'USMS est important, parce qu'il va permet de réorganiser le travail par le passage du système manuel répétitif au system automatique et d'offrir un meilleur service aux utilisateurs par le biais de l'accès à des catalogues distants en line via intranet ou internet et faciliter aux dirigeants la prise de meilleure décision au bon moment.

Pour cela, l'intégration d'un système intégré de gestion de bibliothèque est un programme prioritaire de l'enseignement supérieur, dans le cadre du Maroc numérique, qui a une politique volontariste pour développer l'utilisation des nouvelles technologies de l'information et de la communication dans les établissements d'enseignement supérieur. Ce système d'information

opérationnel permet d'améliorer fortement les services présentés par la bibliothèque traditionnelle.

On peut citer les limitations de la gestion des bibliothèques traditionnelles relevant à l'USMS de Beni Mellal comme suit :

- Problème de réservation en ligne ;
- Le problème de l'ignorance de documents ;
- Problème d'insatisfaction des documents ;
- Manque de la prise de meilleures décisions de l'appart des dirigeants ;
- La lentement des services offerts par les bibliothèques ;
- Problème de partage entre bibliothèque relevant à la même université ;

La figure ci-dessous (cf. figure 57) présente les différentes tâches d'un système informatique de gestion de bibliothèque.



Figure 57 : Management de la bibliothèque

L'auteur [64] définit les objectifs de la gestion des archives documentaires comme suit :

- L'amélioration des services techniques ;

- Le contrôle des coûts (Gestion du budget due aux achats des nouveaux ouvrages) ;
- L'amélioration des services aux usagers ;
- L'amélioration des outils de gestion ;
- La réorganisation de la bibliothèque ;
- La réponse à la crise du système manuel ;
- La coopération entre les différents catalogues et la perspective de l'échange de données bibliographiques.

4.2.2 Indicateurs de performance de la bibliothèque

Dans ce qui suit, on va présenter les différents indicateurs mesurant la performance de la bibliothèque :

- Nombre des inscrits par filières par établissement par date d'adhésion, le nombre des lecteurs ayant l'abonnement expiré. (Figure 61)
- Les documents les plus réservés, les plus prêtés, les moins prêtés qui ne sont jamais prêtés. (Figure 59)
- Savoir les statistiques sur le fond documentaire par types des documents. (Figure 58)
- Les statistiques sur les abonnements (expiré, à renouveler) ;
- Statistiques des exemplaires achetés ou créés (Figure 60).
- L'évolution des dépôts documentaire par année universitaire, par type, par auteurs, etc...
- La surface : combien de mètre carré par nombre de lecteurs ;
- Les collections ;
- Les heures d'ouverture et l'état des locaux ;
- La satisfaction des usagers ;
- La possibilité de réservation à distance ;
- Les formations données ;
- Les systèmes de recommandations, etc...

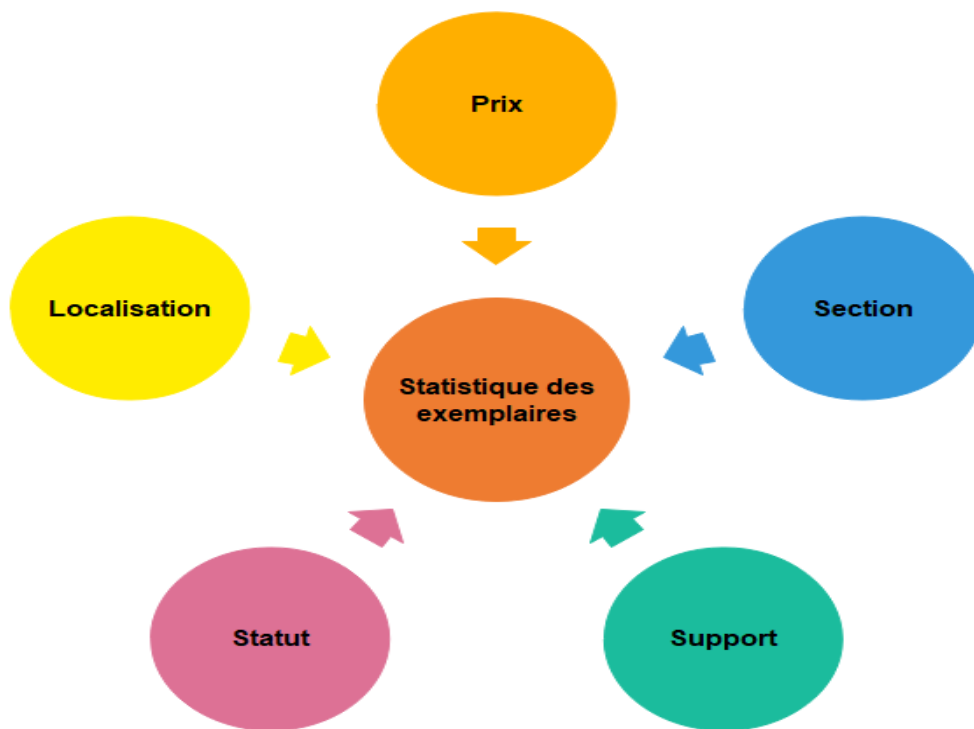


Figure 58 Statistiques des exemplaires

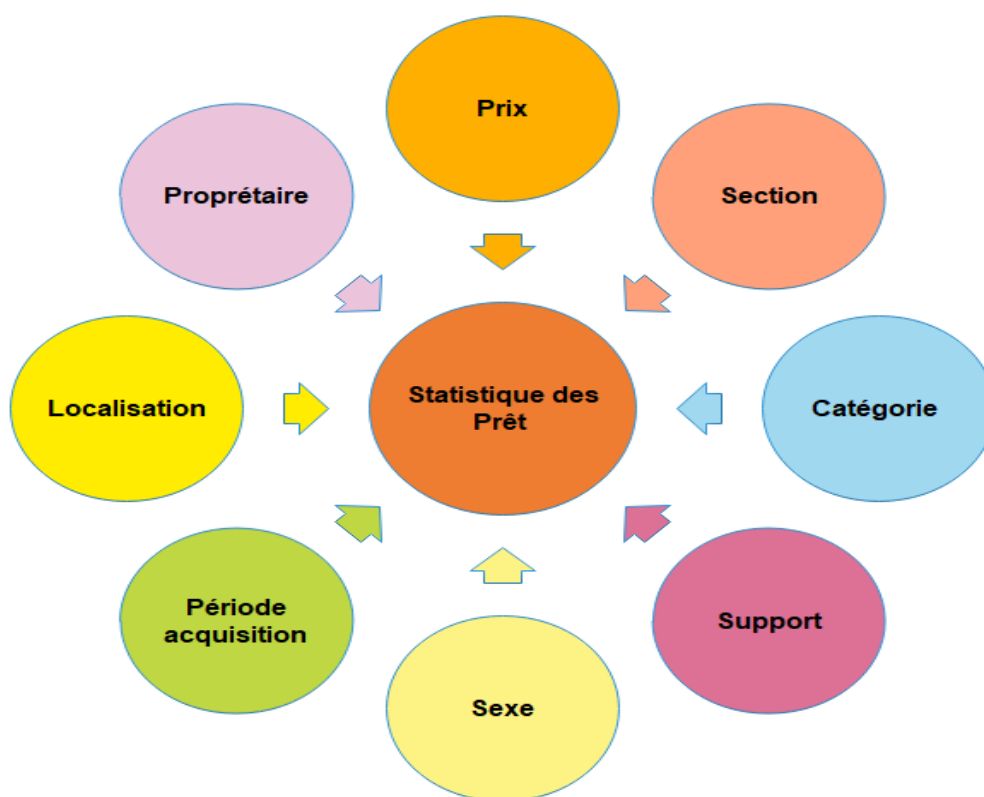


Figure 59 : Statistique de prêt

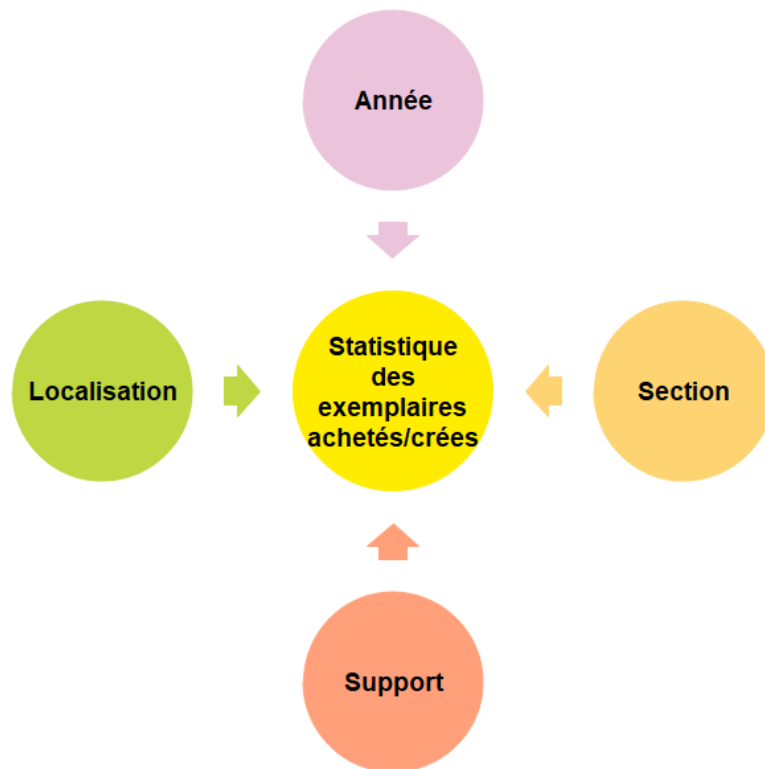


Figure 60 : Statistiques des exemplaires achetés ou cr es

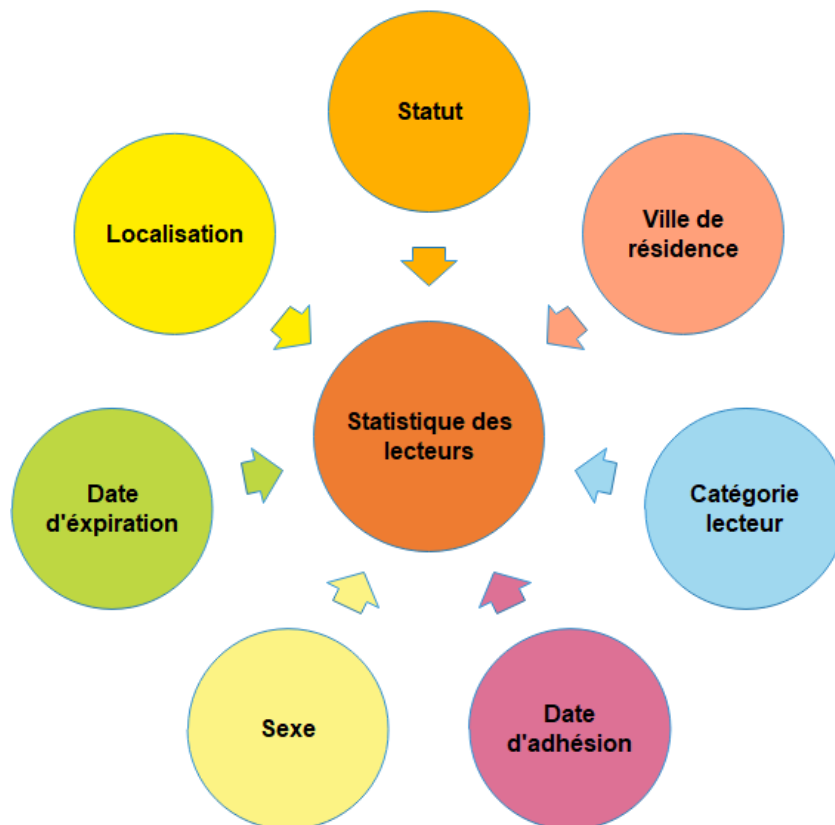


Figure 61 Statistique des lecteurs

Les figures 58 à 61 présentent les différentes statistiques (des exemplaires, de prêt des exemplaires et des lecteurs) selon plusieurs axes de dimension.

4.2.3 Méthode proposée

Nous avons proposé le processus présenté ci-dessous (cf. figure 62) en vue de contribuer à promouvoir la production scientifique locale et de la performance de bibliothéconomie.

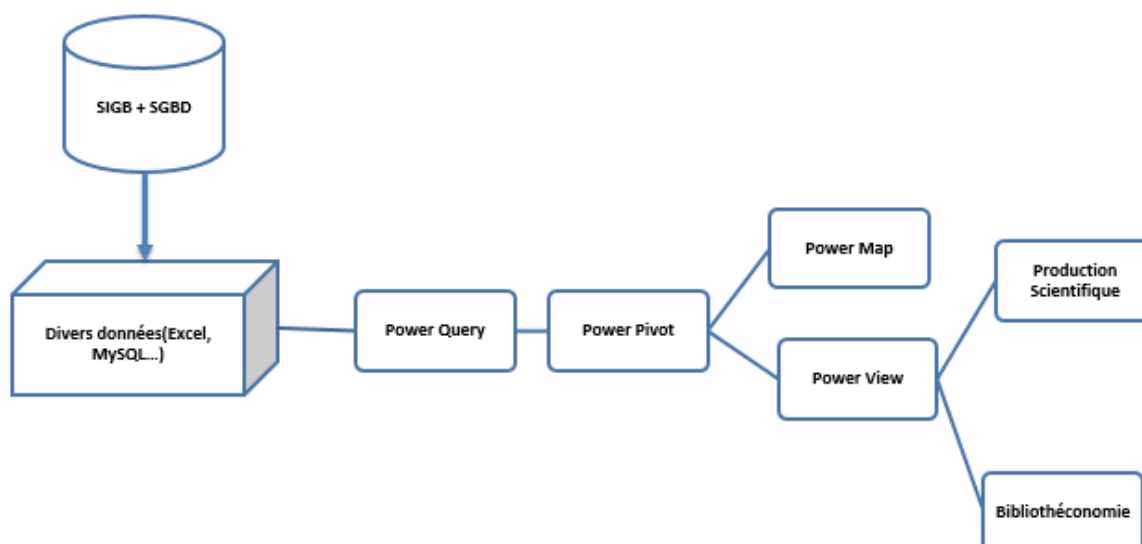


Figure 62 Processus de l'approche proposé

Comme la montre la figure 62, l'approche proposée se compose de plusieurs étapes consécutives. Elle comprend la mise en œuvre du système intégré de gestion de bibliothèque (PHP My Biblio), la collecte, l'extraction, la transformation et la fusion des données avec l'outil de requête Power Query. Après cela, nous utilisons l'outil Power Pivot pour créer un modèle de données définissant les relations entre les données incorporées à la requête, puis nous créons des rapports sous forme de graphiques interactifs utilisant Power View; l'outil de Power Map est utilisée pour présenter les données géo spatiales en 3D. Enfin, ces présentations nous permettent d'évaluer la bibliothéconomie et de savoir l'état actuel des services offerts par la bibliothèque de l'USMS en faveur de ces lecteurs, et son impact sur le rendement universitaire.

L'outil Power designer Bi regroupe les fonctionnalités résumées dans le tableau 18 suivant :

Tableau 18 : Outil power BI

Fonctionnalité	Description
Power Query	- Découvrir les diverses sources de données (Excel, Access, MySQL, Facebook, Wikipédia ...) - Transformer et la fusionner
Power Pivot	- Créer des modèles de données en insérant des relations
Power View	- Créer facilement des rapports et des vues analytiques avec des graphiques interactifs
Power Map	- Explorer des données géo spatiales sur une carte 3D

4.3 Module d'analyse de référentiel de connaissances (Knowledge Management)

4.3.1 Motivation

L'impact des nouvelles technologies d'information et de communication, et de l'orientation du pays vers le Maroc numérique, ont forcé les institutions publiques à examiner la façon dont elles gèrent leur capital intellectuelle (gestion des connaissances), en particulier, dans la gestion de la bibliothèque, qui vise à fournir une gestion solide et durable et de dépasser les limites de ces bibliothèques traditionnelles (l'ignorance et l'insatisfaction des documents).

Le processus de gestion des connaissances est devenu aujourd'hui l'une des préoccupations majeures dans la gestion stratégique de l'organisation. Afin de gérer cette connaissance, il est essentiel de mener à bien les différentes étapes de sa capitalisation comme l'identification, l'extraction et la représentation des connaissances.

Notre objectif est de capitaliser la connaissance des documents d'une bibliothèque publique Marocaine, le cas de l'Université Sultan Moulay Slimane, pour satisfaire les besoins des lecteurs et de contribuer à promouvoir le rendement scolaire.

Dans notre approche, nous présentons une gestion du référentiel des connaissances en utilisant les annotations des documents (Figure 63), afin de renforcer la qualité du service présenté par les bibliothèques, et pour aider à promouvoir la réussite scolaire.

D'après Gottshalk [65], la gestion des connaissances est une méthodologie pour simplifier et améliorer le processus de partage, de distribution, de création, d'acquisition et de compréhension des connaissances de l'entreprise.

4.3.2 Cas d'utilisation de référentielle connaissance

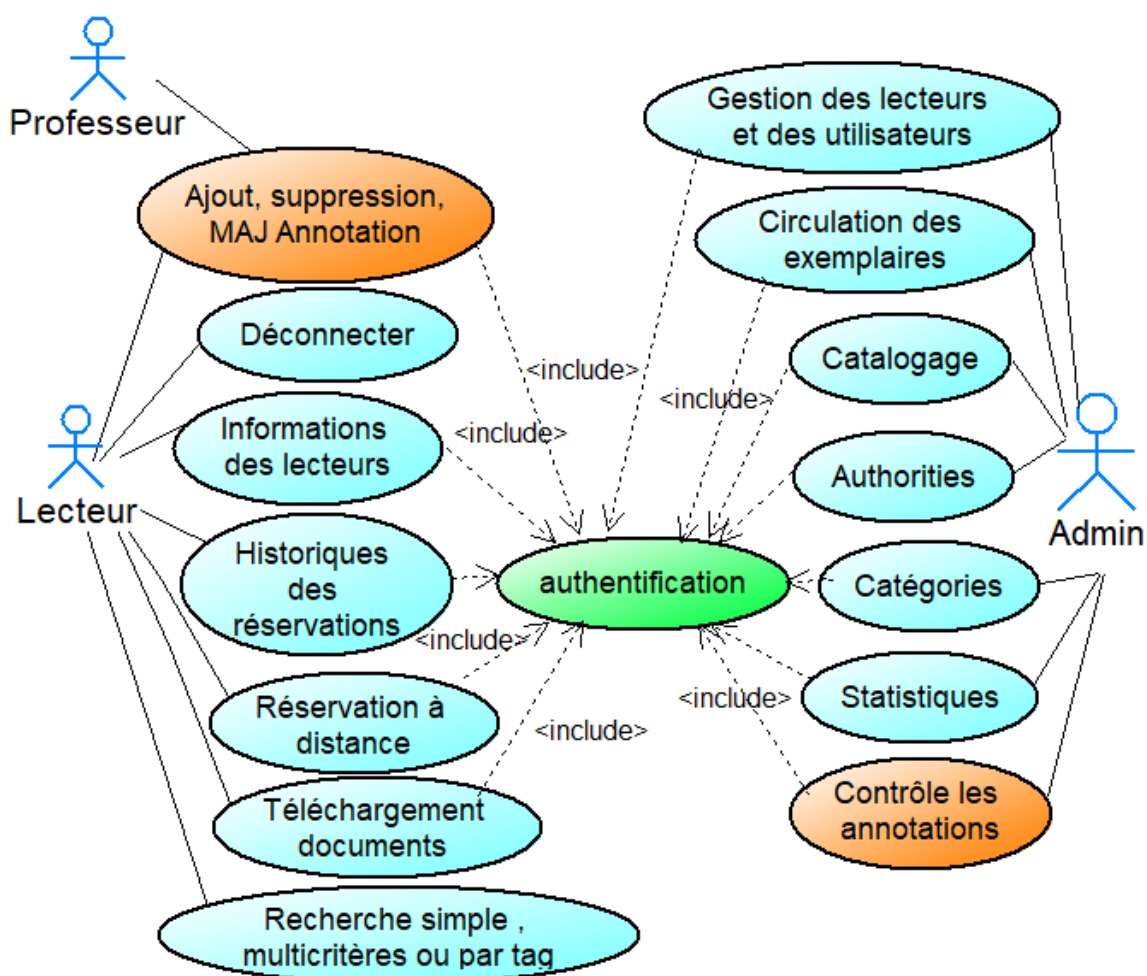


Figure 63 Cas d'utilisation pour annotation des documents

La figure 63 présente le cas d'utilisation du management de la bibliothèque numérique des différents acteurs (Administrateur, professeur et lecteurs) :

Un administrateur peut voir les différentes statistiques, faire le catalogage des notices, gérer la circulation des exemplaires, gérer les lecteurs, les utilisateurs et permettre aussi de contrôler les annotations des autres acteurs.

Les acteurs (professeur et lecteur) peuvent consulter la section OPAC du système intégré de gestion de la bibliothèque (consultation des notices, faire des recherches...), ajoutent des tags et téléchargent des documents numériques...

Le système PMB nous permet d'ajouter des annotations aux exemplaires sous formes des tags. Ces annotations servent comme un moyen de référentiel de connaissances à partir des expériences de lectures.

On peut adopter le modèle de représentation des annotations des notices bibliographiques suivant :

$$D_i = \begin{pmatrix} A_1 \\ \dots \\ A_k \end{pmatrix} \begin{pmatrix} \mathbf{NE} & \mathbf{DD} & \mathbf{CD} & \mathbf{ND} & \mathbf{NS} \\ w_{11} & w_{12} & w_{13} & w_{14} & w_{15} \\ \dots & \dots & \dots & \dots & \dots \\ w_{k1} & w_{k2} & w_{k3} & w_{k4} & w_{k5} \end{pmatrix} \quad (5)$$

Avec :

- Niveau d'étude (NE)
- Domaine de document (DD)
- Catégorie de document (CD)
- Niveau de document (ND)
- Niveau de satisfaction (NS)
- w_{kj} est l'intersection de l'annotation A_i et le descripteur représente le poids du descripteur dans l'Annotation A_i dans le document (notice) D_i .

Ce système proposé permet en premier lieu de catégoriser les notices bibliographiques de la bibliothèque de l'USMS selon les descripteurs indiquées ci-dessus (NE, DD, CD, ND et NS) en utilisant les méthodes de classification automatiques. En deuxième lieu ce système permet aux lecteurs d'affiner la recherche en se basant sur les cinq descripteurs.

4.4 Module d'analyse des sentiments/opinions

4.4.1 Motivations

Récemment, les médias sociaux sont apparus non seulement comme un moyen de communication personnel, mais aussi comme un moyen de communiquer des opinions sur les produits et services ou même les événements politiques. En raison de sa popularité, il y a une quantité massive d'avis ou d'opinions d'utilisateurs produites et partagées quotidiennement. Twitter, Facebook sont parmi les sites des réseaux sociaux les plus utilisés. L'extraction des opinions des utilisateurs à partir des données des réseaux sociaux n'est pas une tâche simple ; cela peut être accompli de différentes manières. Dans ce travail, une approche open source est présentée, dans laquelle les données de microblogs twitter ont été collectées, prétraitées, analysées et visualisées à l'aide d'outils open source pour analyser le contenu et analyser les sentiments des utilisateurs.

4.4.2 Méthodes utilisées pour analyser les sentiments : cas des étudiants

a) En utilisant les nuages de mots

La figure 64 présente le processus de génération de nuage de mots en utilisant une base de données d'enquête et des packages du logiciel R.

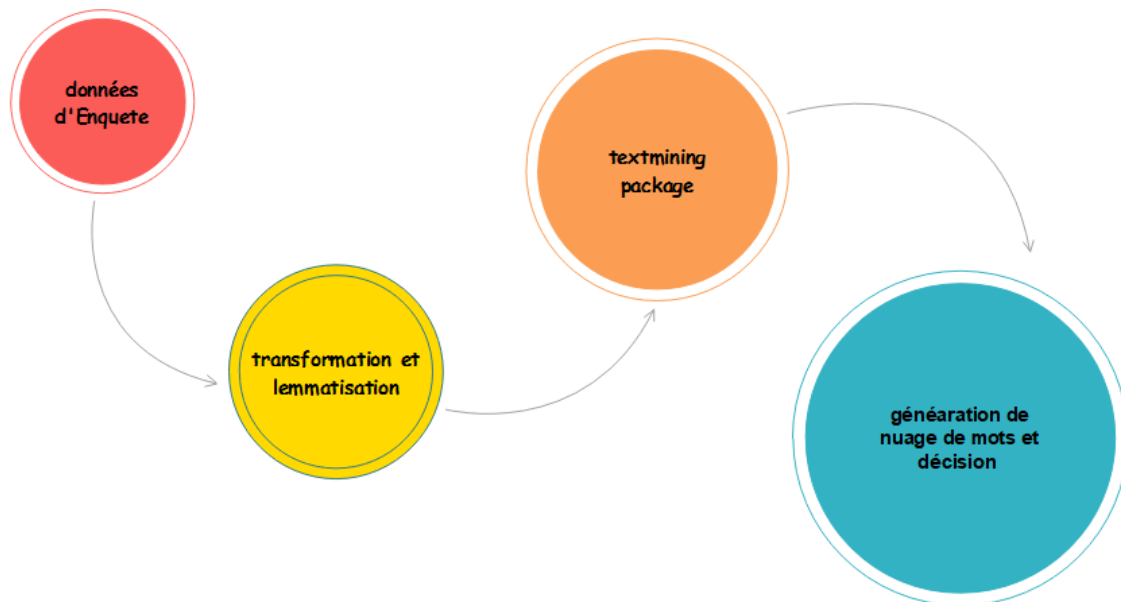


Figure 64 Processus de la génération du nuage de mots des données extraites d'une enquête

Dans cette étude, nous avons installé et chargé les packages suivants :

- Package d'exploration de texte (txetmining) : tm [66]

Ce package permet d'analyser des corpus de documents et de textes. Nous utilisons la commande suivante pour installer ce package :

```
install.packages ("tm")
```

- Package de Stemming : SnowballC [67]

Ce package permet d'implémenter l'algorithme de Stemming de Porter en vue de réduire les mots à une racine commune pour faciliter la comparaison du vocabulaire. Ce package supporte plusieurs langues.

Nous adoptons la commande suivante pour installer ce package :

```
install.packages ("SnowballC")
```

- Génération de nuages de mots : wordcloud [68]

Le package permet de générer le nuage de mots. Pour l'utiliser, nous devons l'installer avec la commande suivante : `install.packages ("wordcloud")`

- Palette de couleurs : RcolorBrewer [69]

Ce paquet peut créer de belles palettes de couleurs en particulier pour les cartes thématiques. Nous utilisons la commande suivante pour installer ce package :

```
install.packages ("RColorBrewer")
```

b) **Analyse d'opinion en se basant sur le lexique des opinions**

On peut utiliser des corpus des lexiques d'opinions (Sentiwordnet) [70] en vue de déterminer la polarité d'un nouveau texte ou d'opinion. Chaque mot du lexique SentiWordNet a trois valeurs entre 0 et 1 : positif, négatif et objectif dont la somme de ces scores est égale à 1, les mots dont les scores positifs et négatifs sont faibles ont très probablement un sentiment neutre.

4.5 Approche proposée pour l'intégration d'un Progiciel de gestion intégré ERP dans le management de l'université

4.5.1 Motivation et objectifs

D'après l'analyse de l'état de l'art sur les systèmes intégrés de gestion, notamment notre recensement sur ceux qui sont intégrés au sein de notre établissement d'étude. Nous avons

déduit que notre établissement n'adopte aucun ERP à la fin de l'année 2012, qui permet de réaliser le management de la majorité des transactions effectuées au sein de l'établissement. Donc nous sommes motivés de suivre la méthode suivante (figure 65) en vue de bien choisir l'ERP qui s'adapte mieux à l'établissement en terme de besoin, de coût et de qualité.

4.5.2 Méthode de mise en œuvre de l'ERP

Nous avons utilisé le processus qui est schématisé dans la figure 65 en vue de mettre en œuvre notre ERP fiable à moindre coût.

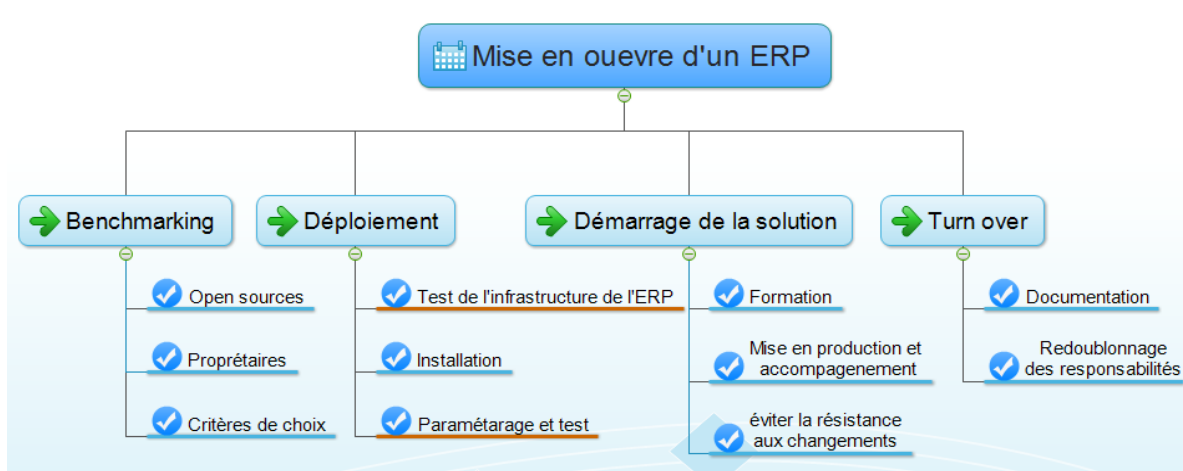


Figure 65 Processus de mettre en œuvre notre ERP

4.6 Approche proposée pour l'analyse de données massives de l'université

4.6.1 Motivation et objectifs

D'après l'analyse des besoins, le recensement des différents systèmes intégrés au sein de l'établissement d'étude, l'état de l'art sur les différentes méthodes d'analyse de données, l'étude de l'état de l'art sur la notion de Big Data, nous avons confirmé qu'il est nécessaire de penser à intégrer des nouveaux systèmes d'analyse des données massives structurées, semi et non structurées.

4.6.2 Méthode de mise en œuvre d'un Système intégré pour le Big Data

Après avoir recensé les différents types de sources d'information de l'Université, maintenant nous allons présenter notre modèle d'analyse des données caractérisé par les 10Vs.

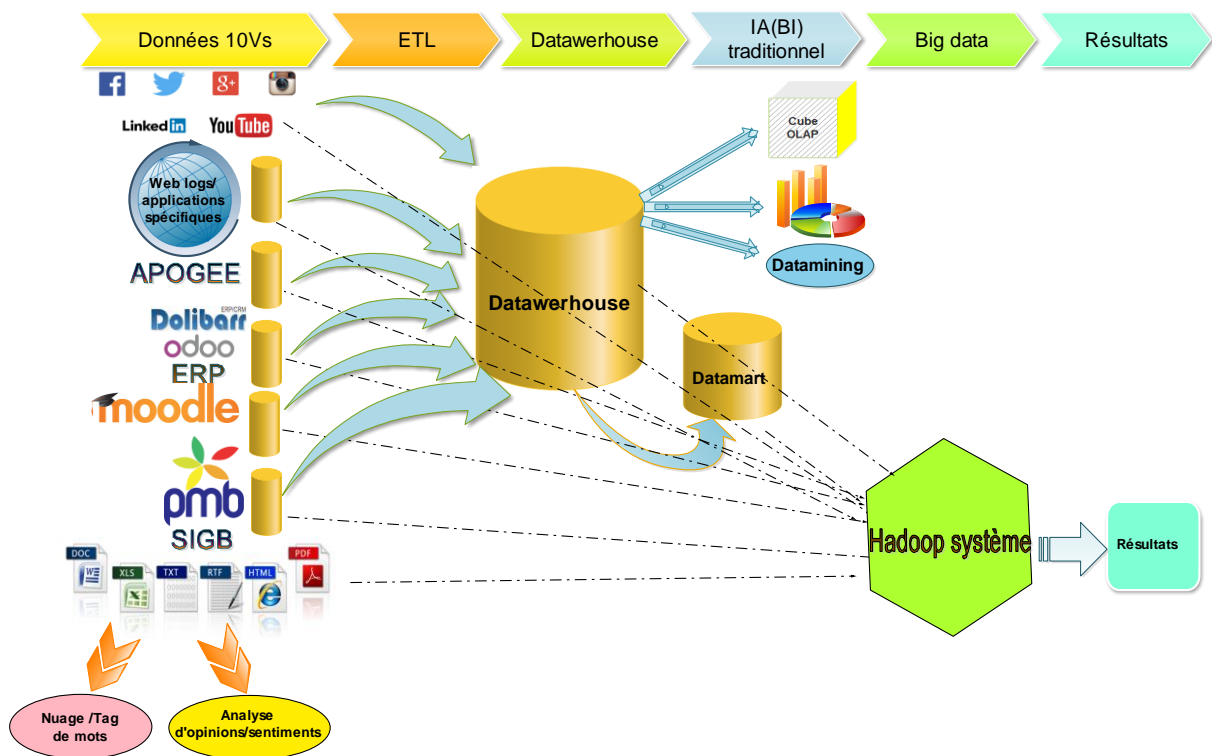


Figure 66 Schéma global du projet

4.7 Conclusion

Dans ce chapitre, nous avons entamé la description de nos contributions dans le management de l'université. En premier lieu, nous avons présenté l'approche de gestion de la bibliothèque universitaire à travers l'instauration du système intègre PMB, ensuite, nous avons parlé sur l'utilisation de ce dernier SIGB pour le référentiel de connaissances par le biais des annotations des ouvrages ou bien des notices bibliographiques. Une contribution d'analyse d'opinion ou de sentiment a été proposée. Puis nous avons présenté une approche qui s'intéresse à l'intégration d'un Progiciel de gestion intégré ERP dans le management de l'université. Et finalement nous avons proposé une méthode d'analyse de données massives des différents systèmes intégrés au sein de l'Université.

Dans ce qui suit, on abordera l'implémentation et la mise en œuvre des différentes contributions ainsi qu'on présentera quelques résultats retournés via les outils utilisés.

Chapitre 5. Implémentation et résultats

Chapitre 5. Implémentation et résultats

5.1 Introduction

Dans ce chapitre, on va présenter la démarche et les résultats de développements des deux modules de notre application.

5.2 Réalisation du module SIGB (PMB)

5.2.1 Acquisition et installation des logiciels de management de la bibliothèque

❖ *Étape 1 : Acquisition des logiciels*

Pour mieux tourner des fonctionnalités spécifiques de PMB sur un serveur local, il est requiert d'installer quelques logiciels spécifiques et de configurer certains fichiers.

La réalisation du module PMB nécessite les logiciels suivants :

- PMB version 5.0.4 ;
- Logiciel de lecteur de code barre ;
- EasyPhp qui comprend le trio (apache web serveur, MySQL et PHP).

❖ *Étape 2 : Procédure d'installation en local*

Après avoir téléchargé EasyPhp à partir du site officiel, on commence à installer EasyPhp. Ensuite, on modifie les fichiers de configuration php.ini et my.ini (Voir annexe 4)

- *Téléchargement et installation de la nouvelle version de pmb*

Avant de commencer à télécharger la version de PMB, il faut avoir une idée sur la version de PHP et MySQL installée selon les critères définie dans l'annexe 5.

On télécharge la version pmb 5.0.4 à partir du site web des différentes versions de pmb service, puis on suit les étapes suivantes :

❖ *Étape 3 : Décompresser le fichier zip de PMB dans le répertoire adéquat*

❖ *Étape 4 : Paramétrage du système et de la base de données PMB (Annexe 5)*

❖ *Étape 5 : Création des utilisateurs de l'application et attribution des droits d'accès ;*

Comme chaque système intégré, le PMB permet de gérer les utilisateurs en les attribuant les droits d'accès.

❖ **Étape 6 : Importation des lecteurs et des exemplaires ;**

Cette étape est primordiale et qui se répète pour chaque année Universitaire. La figure 79 (Annexe 6) illustre le modèle des notices bibliographiques et la conversion du fichier .csv en format UNIMARC.

❖ **Configuration du portail OPAC ;**

Le SIGB est un système paramétrable. Dans la partie administration, on peut configurer l'OPAC de notre bibliothèque Universitaire.

❖ **Étape 7 : Teste de la Recherche des notices bibliographiques ;**

Dans la partie OPAC, on peut faire des recherches simples ou des recherches multicritères.

❖ **Étape 8 : Gestion électronique des documents numériques sous PMB ;**

PMB permet de gérer des documents numériques.

❖ **Étape 9 : Procédure d'installation sur un serveur distant**

De la même manière d'installer le PMB en local, se réalise en serveur sauf qu'il fallait configurer le serveur de la base de données.

5.2.2 Évaluation de l'activité en bibliothèque de l'USMS de Beni Mellal en utilisant le SIGB PMB et l'outil décisionnel power designer BI

Cette section de notre travail sera consacrée à l'évaluation des bibliothèques relevant à l'université Sultan Moulay Slimane. Notamment la bibliothèque de la FP qui a pris l'initiation d'adopter ce système informatique.

La figure 67 présente la page d'authentification au Système Intégré de Gestion de Bibliothèque PMB en se connectant sur le serveur de l'établissement de l'USMS.

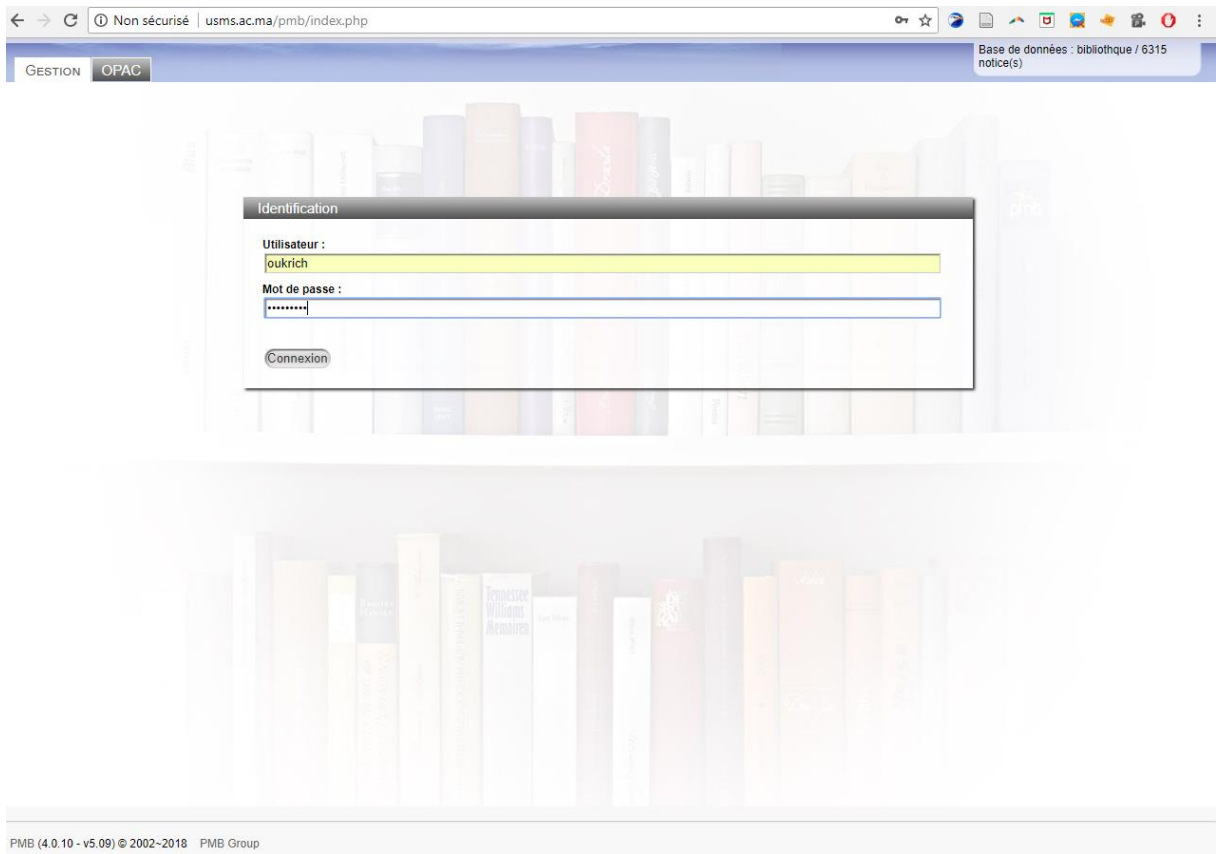


Figure 67 Fenêtre de connexion au système PMB

La figure 68 et le tableau 19 présentent des statistiques du nombre de prêts des exemplaires durant la période entre 2013 et 2018.

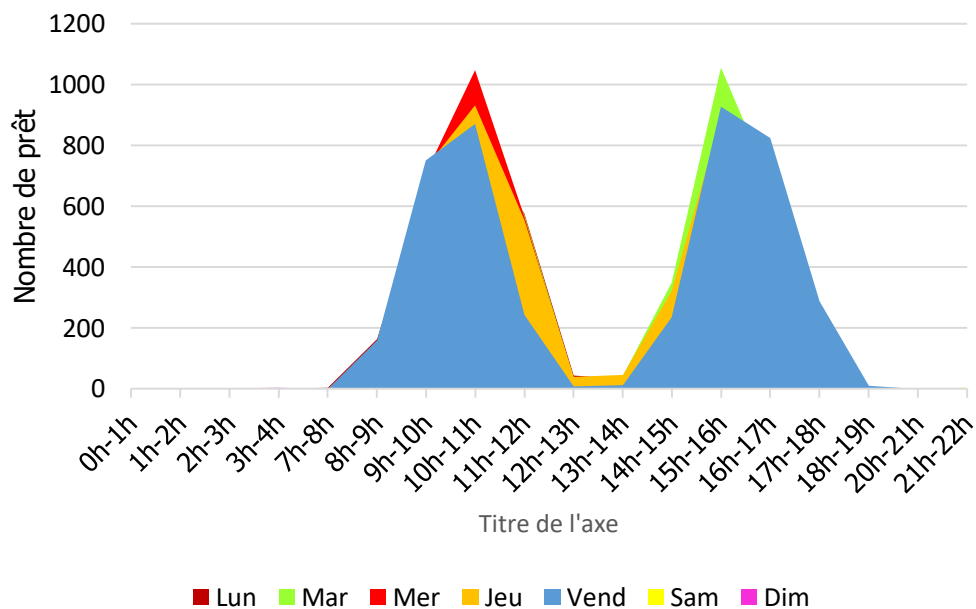


Figure 68 Nombre de prêts par jour et par tranche horaire de la bibliothèque FP

Tableau 19 Nombre de prêt par année et par mois entre 2013 et 2018

Mois Année	1	2	3	4	5	6	7	9	10	11	12	Total général /année
2013				1	34	7			1	789	753	1585
2014	403	98	733	1029	1076	169	13	10	669	603	1217	6020*
2015	583	76	356	795	764	337	37	6	378	739	752	4823
2016	111	38	123	582	718	162	44	10	247	732	591	3358
2017	109	53	716	725	718	72	41	10	182	858	631	4115
2018	65	97	716	522	365	65	41	18	407	0	359	2655
Total général	1271	362	2644	3654*	3675*	812	176	54	1884	3721*	4303*	22556

D'après la figure 68 et le tableau 19, on déduit que le maximum de prêt des exemplaires se fait généralement dans les trois premiers jours de la semaine (Lundi, Mardi et Mercredi) et dans les tranches horaires entre 10h à 11h du matin et entre 15h à 16h après midi. Aussi les prêts s'effectuent durant la période des examens dans les mois de Novembre, de Décembre, d'Avril et de Mai.

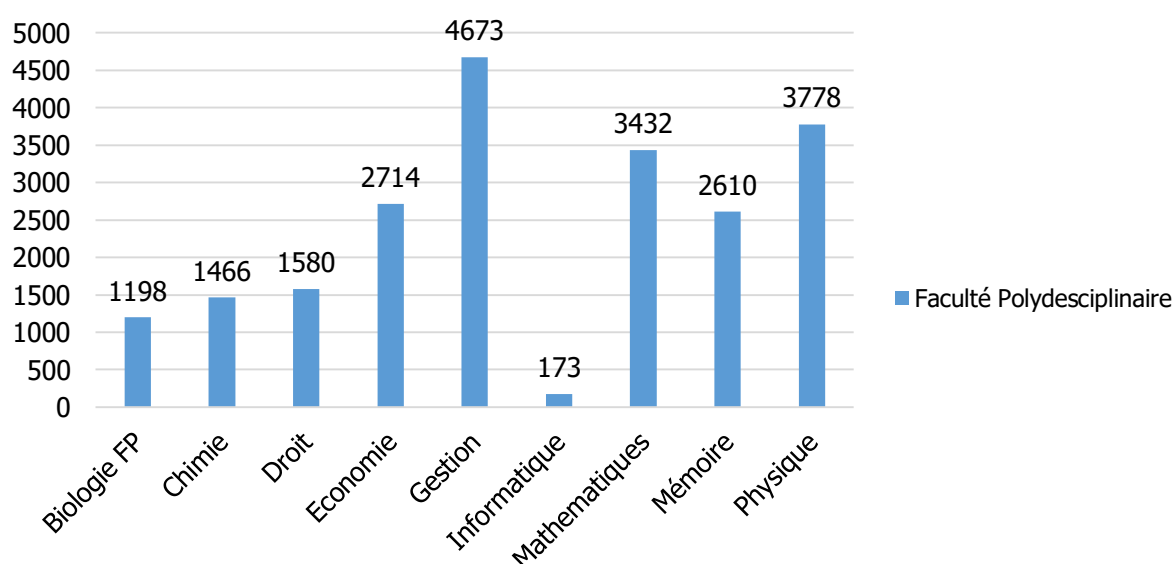


Figure 69 Nombre total des prêts par catégorie (section) depuis 1 septembre 2013 au 1 Juin 2018

La figure 69 présente le nombre total de prêt des exemplaires par section ou catégorie, on remarque que la section la plus demandée est de gestion en raison du nombre important des étudiants inscrits dans la filière sciences économique et de gestion ou en droit.

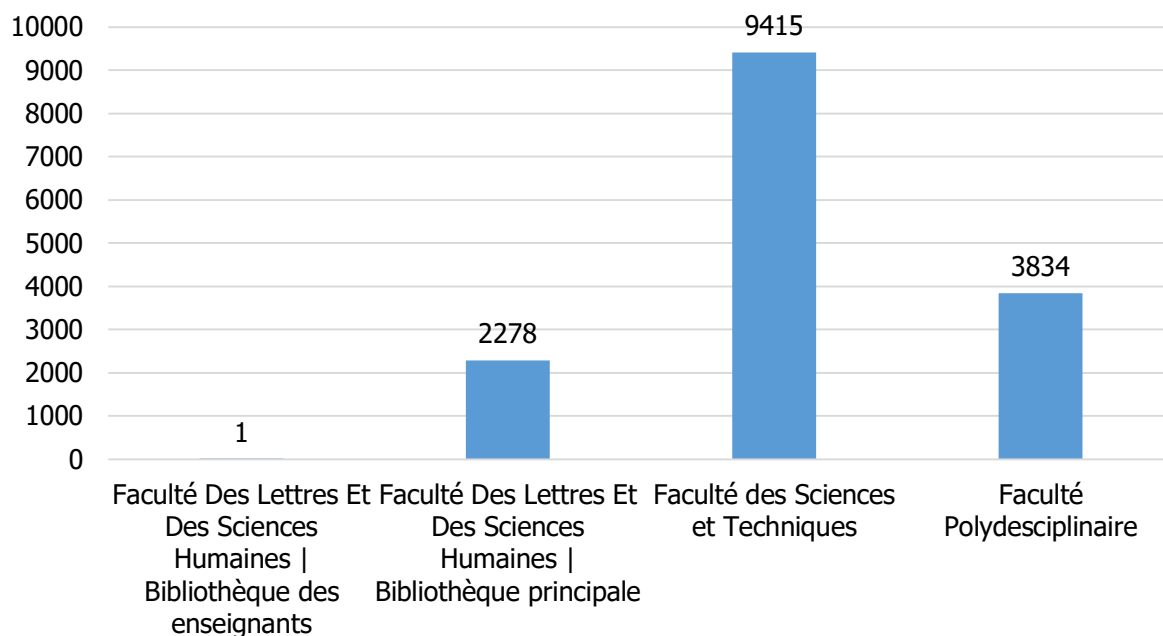


Figure 70 Nombre des exemplaires par localisation (Consulté le 22 Mars 2019)

La figure 70 illustre le nombre total des exemplaires par établissement dans le système PMB.

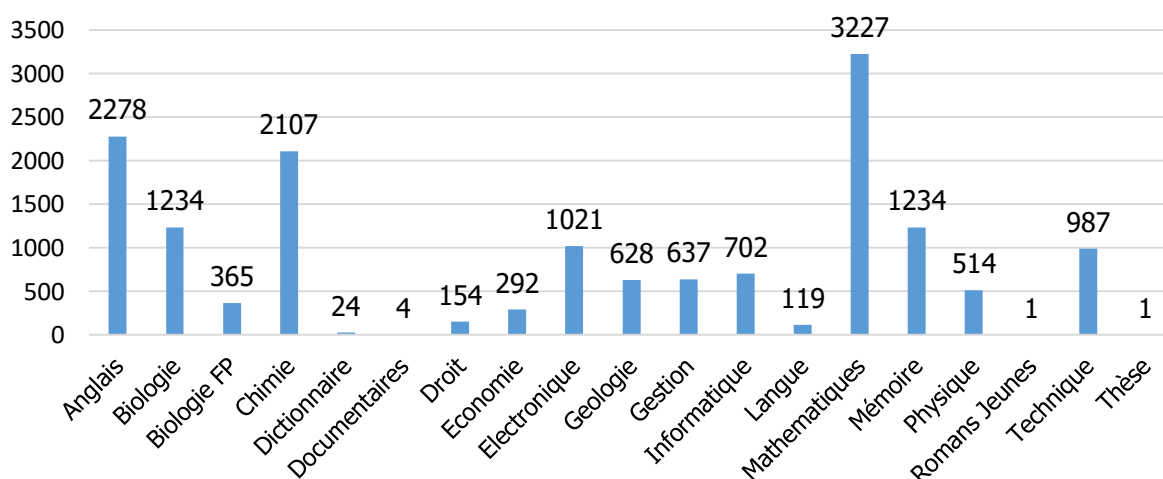


Figure 71 Nombre d'exemplaire par section (le 22 Mars 2019)

La section mathématique occupe la première place du nombre des exemplaires (cf. figure 71)

Le tableau 20 présente la liste des 40 documents les plus prêtés dans la bibliothèque numérique de la FP de Beni Mellal depuis le 1 avril 2013 jusqu'au le 01 Juin 2018.

Tableau 20 Notices bibliographiques par ordre décroissant de nombre de fois prêtés

Titre	Nombre de fois prêtés
economie politique: exercices +examens corriges	388
gestion des ressources humaines	247
economie monetaire	245
gestion financiere manuel + applications	241
comptabilite analytique de gestion -5eme edition-	233
td electronique –copie	233
statistique descriptive examens + corriges	230
campus mini manuel de probabilites statistique	222
statistiques pour l'economie et la gestion (traduction de la 4 eme edition americaine)	215
mini manuel de mecanique du point	211
fiscalite marocaine de l'entreprise	210
economie politique –copie	194
analyse 1re annee - 2eme edition - exercices corriges	190
mathematique pour l'economie analyse-algebre	190
droit des affaires	162
economie monetaire et financiere	158
campus mini-manuel de chimie organique	154
Microeconomie	149
exercices de gestion financiere finance d'entreprise en 26 fiches 3e edition	149
Electromagnetisme	148
mini manuel d'algebre - cours et exercices corriges	148
comptabilite generale + gestion des entreprises	145
campus mini manuel de comptabilite	145
introduction à la gestion des ressources humaines	145
campus exercices de gestion financiere finance d'entreprise en 26 fiches 2e edition	143
campus physique tout-en-un pour la licence - cours, applications et exercices corriges	141
maths mpsi pcsi ptsi - 1re annee - exercices et problemes	139
algebre et geometrie euclidienne 1re annee pcsi / ptsi - cours avec exercices corriges	134
math 1re annee mpsi - h prepa tout-en-un	133
optique geometrique	130
analyse financiere, aide-memoire -3eme edition	130
la monnaie - 2e edition	128
comptabilite analytique de gestion	127
chimie pcsi - 1ere annee - exercices et problemes	124
h-prepa - physique - 1re annee mpsi/pcsi/ptsi - exercices et problemes - lmd	123
td - algebre - 3e edition - travaux diiriges	123
microeconomie: cours,methodes,exercices corrigees	122
physique generale 2 electricite et magnetisme	119
algebre lineaire - 3e edition	118
comptabilite de gestion - 4eme edition - en 20 fiches	118

Tableau 21 Exemples des documents qui ne sont jamais prêtés

Propriétaire	Localisation	Section	Statut	Cote	Titre	Auteur principal	Editeur	Année d'édition	Code-barres
Fonds propre	Faculté Polydisciplinaire	Physique	Document en bon état	P69-1	CAMPUS PRINCIPES DES CIRCUITS ELECTRIQUES	DIXNEUF+BELLOUVET	Dunod		FP-818
Fonds propre	Faculté Polydisciplinaire	Physique	Document en bon état	P69-2	CAMPUS PRINCIPES DES CIRCUITS ELECTRIQUES	DIXNEUF+BELLOUVET	Dunod		FP-817
Fonds propre	Faculté Polydisciplinaire	Physique	Document en bon état	P69-3	CAMPUS PRINCIPES DES CIRCUITS ELECTRIQUES	DIXNEUF+BELLOUVET	Dunod		FP-816
Fonds propre	Faculté Polydisciplinaire	Physique	Document en bon état	P69-4	CAMPUS PRINCIPES DES CIRCUITS ELECTRIQUES	DIXNEUF+BELLOUVET	Dunod		FP-815
Fonds propre	Faculté Polydisciplinaire	Physique	Document en bon état	P69-5	CAMPUS PRINCIPES DES CIRCUITS ELECTRIQUES	DIXNEUF+BELLOUVET	Dunod		FP-2814
Fonds propre	Faculté Polydisciplinaire	Physique	Document en bon état	P74-2	CHAINES DE MARKOV - COURS ET EXERCICES CORRIGES	GRAHAM GRAHAM-C	Dunod		FP-913
Fonds propre	Faculté Polydisciplinaire	Physique	Document en bon état	P74-4	CHAINES DE MARKOV - COURS ET EXERCICES CORRIGES	GRAHAM GRAHAM-C	Dunod		FP-911
Fonds propre	Faculté Polydisciplinaire	Physique	Document en bon état	P74-5	CHAINES DE MARKOV - COURS ET EXERCICES CORRIGES	GRAHAM GRAHAM-C	Dunod		FP-914
Fonds propre	Faculté Polydisciplinaire	Physique	Document en bon état	P75-1	CIRCUITS ELECTRIQUES	MAHMOOD NAHVI	Ediscience		FP-1056
Fonds propre	Faculté Polydisciplinaire	Physique	Document en bon état	P75-2	CIRCUITS ELECTRIQUES	MAHMOOD NAHVI	Ediscience		FP-910
Fonds propre	Faculté Polydisciplinaire	Physique	Document en bon état	P75-4	CIRCUITS ELECTRIQUES	MAHMOOD NAHVI	Ediscience		FP-909

Nous avons 693 exemplaires qui n'ont jamais prêtés depuis l'instauration du système pmb jusqu'au le 1 juin 2018. (Tableau 21)

Dans ce travail, nous avons présenté la gestion des ressources d'information de la bibliothèque publique marocaine, le cas de l'USMS de Beni Mellal, en utilisant les outils de l'intelligence artificielle BI, nous avons utilisé la bibliothèque de la Faculté Polydisciplinaire comme une bibliothèque de test. Nous avons trouvé que la bibliothèque souffre de nombreux problèmes tels que l'ignorance et l'insatisfaction des documents et les problèmes de manque de décision. Dans notre étude, nous avons constaté que les documents numériques sont très peu ou pas disponibles dans les bibliothèques publiques d'USMS. Dans la suite nous allons proposer une nouvelle méthode qui utilisera un algorithme pour aider à minimiser les documents les plus ignorés, et d'optimiser la recherche de la bonne information à la bonne personne au bon moment.

5.3 Application de la référentielle connaissance

Nous avons utilisé la méthode de nuage de mots pour faciliter la recherche de documents dans le système de gestion de bibliothèque intégrée (PMB) par des mots les plus fréquents.

Dans cet article, nous présenterons le mot nuages de différentes thèses, dissertations de fin d'étude existant dans deux bibliothèques (bibliothèque faculté polydisciplinaire et bibliothèque de département de sciences de la terre de la faculté des sciences et techniques de Beni Mellal).

Recommandation :

- Pour les étudiants, ils peuvent affiner leurs recherches dans le système de gestion de bibliothèque intégré PMB en utilisant les mots les plus fréquents (ex: pour le cas de la bibliothèque de la FP on peut rechercher par les mots Maroc, management, entreprise, banque ...)

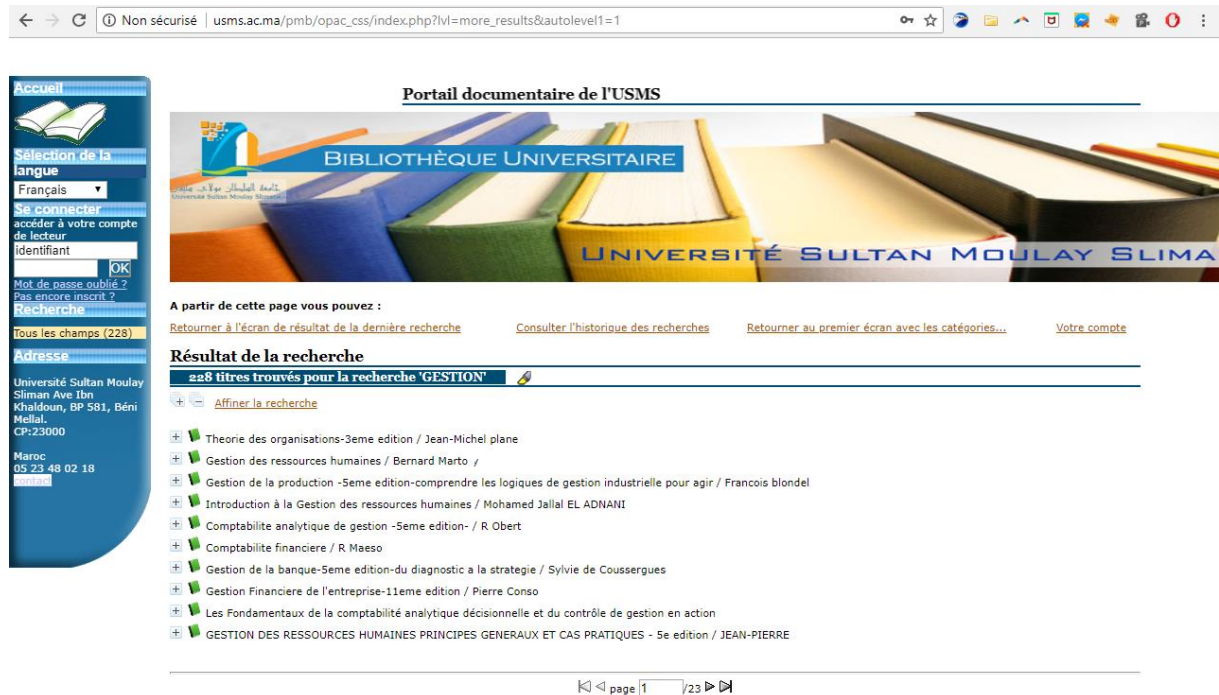


Figure 75 Exemple d'utilisation des mots les plus fréquents

- Pour les enseignants, il est le temps de réfléchir à de nouveaux cas d'étude au lieu de poser des problématiques ou des sujets déjà traitées et consommées en analysant ces nuages de mots ou bien rediriger les étudiants à consulter ces références bibliographiques locaux.

5.4 Réalisation de l'ERP de l'USMS

Nous avons adopté le système de gestion intégré Dolibarr pour les raisons qu'on nous avons déjà citées dans la section de Benchmarking des outillages. Nous avons implémenté la dernière version doliwamp7.2.

Le Doliwamp c'est l'auto installateur pour Windows, facile à installer, plus sécurisé et simple à développer en les ajoutant des modules de notre choix.

Le système intégré Doliwamp nous permet de gérer toutes les transactions de chaque faculté : la Gestion des ressources humaines, la chaîne logistique, la gestion des fournisseurs et des clients, la gestion de paiement, la gestion électronique des documents...

La figure suivante présente le tableau de bord de notre système intégré :

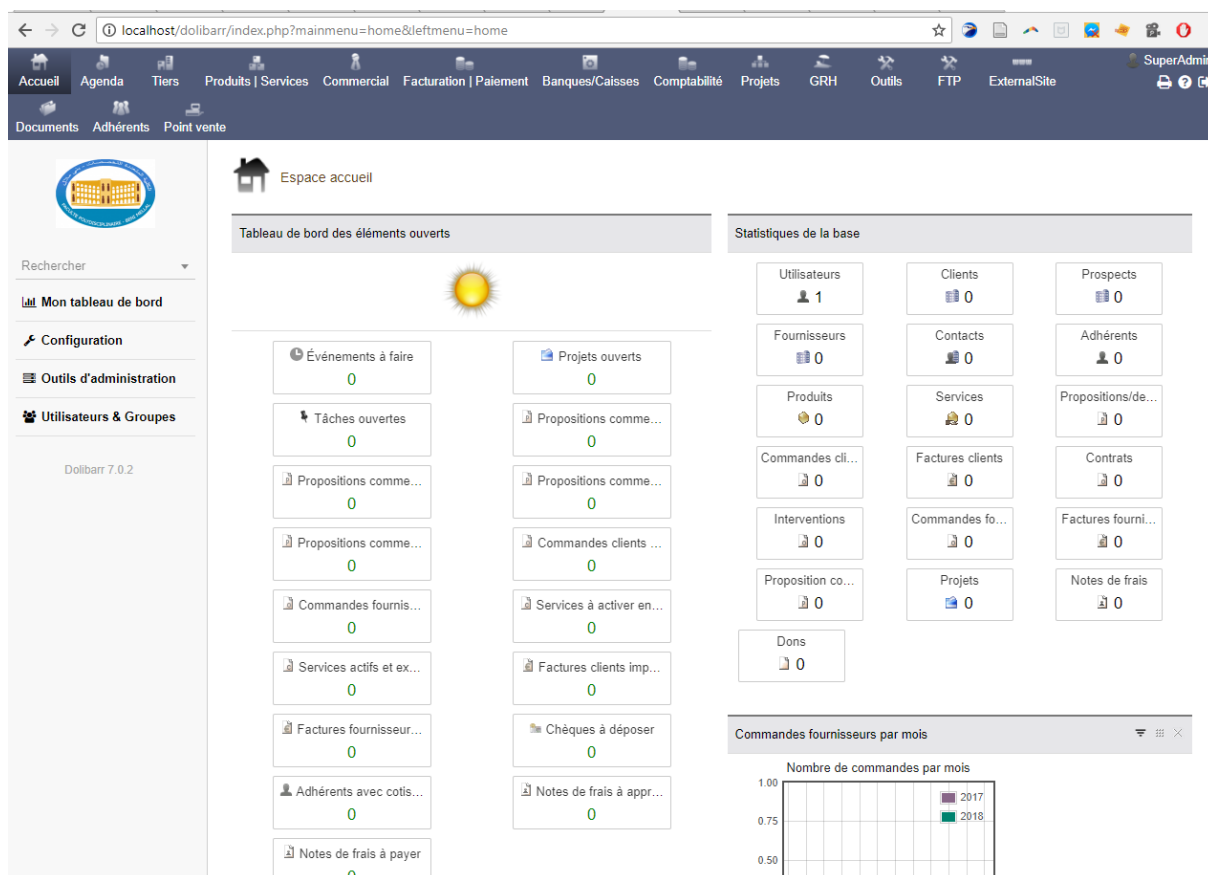


Figure 76 Tableau de bord du progiciel ERP proposé pour FP de Beni Mellal

5.5 Réalisation du module Analyse des sentiments

En utilisant le logiciel R, nous avons trouvé le résultat présenté dans la figure suivante :

- En se basant sur le lexique des opinions

Pour analyser les opinions des différents enquêtés, nous avons utilisé les packages du logiciel R suivants :

- Library ("Tm") [66]
- Library("SnowballC") [67]
- library("wordcloud") [68]

- library("RColorBrewer") [69]
- library(ggplot2) [71]
- library(lubridate) [72]
- library(Scale) [73]
- library(stringr) [74]
- library(syuzhet) [75]

Cette analyse nous a permis de partager le lexique de notre corpus (enquête) en des degrés de sentiments (en colère, anticipation, dégoût, peur, négative, positive, tristesse, étonnement et de confiance).

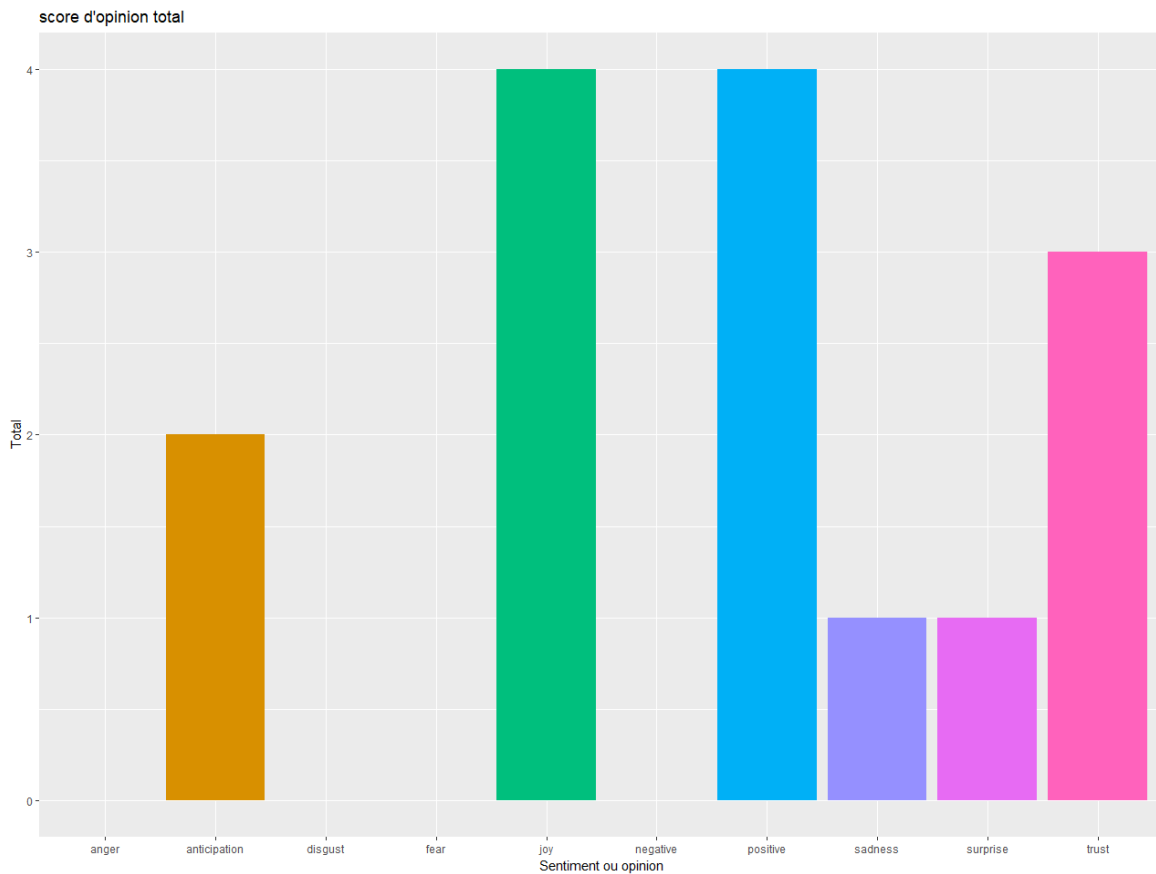


Figure 77 Score des opinions analysées d'une enquête de satisfaction

Ce résultat nous montre que les étudiants ne sont pas fortement satisfaits de tous les services rendus par la bibliothèque traditionnelle.

5.6 Conclusion

Nous avons présenté dans ce chapitre une vue pratique des approches proposées pour le management des documents de l'université. Tout d'abord nous avons évalué quantitativement la bibliothèque de l'université via l'utilisation des Système intégré de Gestion de Bibliothèque PMB. Ensuite nous avons contribué à mesurer le taux de satisfaction des lecteurs de la bibliothèque en adoptant l'outil d'analyse d'opinion. Puis nous avons utilisé le PMB comme une source de référentielle de connaissances. Ensuite, nous avons implémenté le système Dolibarr pour gérer les transactions de l'établissement. Enfin nous avons présenté l'architecture globale d'utilisation d'un système d'analyse de données massives à savoir Hadoop HDFS pour analyser les différentes sources d'information dans l'université.

Conclusion générale et perspectives

Dans cette thèse, nous nous sommes intéressés à l'étude de la modélisation de la gestion des documents dans le cas d'un établissement universitaire public, en particulier cas de l'USMS de Beni Mellal. Nous avons suivi une méthodologie de réalisation d'un tel projet. Tout d'abord, nous avons analysé le besoin en termes de progiciels et en niveau de satisfaction des différents acteurs de l'Université. Puis, nous avons planifié le projet en vue d'estimer le coût et la durée du projet. Ensuite, nous avons proposé une étude de Benchmarking en vue de sélectionner les progiciels open sources les plus adaptés à l'environnement techniques, matériels et compétences de notre Université. Nous avons choisi le SIGB (PMB) dont lequel nous avons montré qu'il permet d'offrir une meilleure qualité de service aux lecteurs tout en réduisant le temps d'accès à l'information, une réservation à distance... ainsi pour le décideur de bien prendre les bons décisions dans les bons moments. Le choix de l'ERP Dolibarr va permettre à leur tour de bien gérer toutes les transactions de l'Université, notamment la gestion électronique des documents, la Gestion des Ressources Humaines, la gestion de la chaîne logistique et ainsi qu'il est facile à ajouter d'autres modules à ce progiciel selon le besoin de l'établissement.

L'utilisation des nuages de mots nous a permis de bien analyser les enquêtes ainsi que la possibilité de l'utiliser pour présenter visuellement un corpus de mémoires ou de thèses pour savoir les sujets qui sont déjà traités et consommés ou pour affiner la recherche des notices bibliographiques de la bibliothèque universitaire.

Notre mémoire a été subdivisé en 5 chapitres. Dans le premier chapitre, nous avons élaboré une méthodologie d'analyse de l'existant et des besoins, la présentation du cadre institutionnel de l'étude et la planification du projet. Ensuite, le deuxième chapitre est dédié à présenter l'état de l'art sur les Systèmes d'Information ainsi que les différentes méthodes d'analyse de données. Le troisième chapitre est consacré à proposer une démarche détaillée de l'étude de la sélection des outillages open source à savoir le SIGB, un ERP fiable et robuste et un système d'analyse de données massives les plus adaptés possible à l'université ainsi que nous avons proposé une description des méthodes d'analyse d'opinion. Le quatrième et le cinquième chapitre sont dédiés pour la présentation des contributions et leurs implémentations.

Au terme de ce mémoire, nous avons validé les résultats suivants :

- L'instauration du SIGB au sein de l'établissement de test (FP de Beni Mellal) a contribué efficacement dans le management de la bibliothèque (prendre des meilleures décisions, facilite la recherche documentaire, savoir la disponibilité des exemplaires à distance via internet, possibilité de réserver à distance, savoir les documents les plus, les moins ou les jamais prêtés...)
- L'outil d'analyse d'opinion permet d'analyser le niveau de satisfaction des lecteurs de la bibliothèque ;
- Utilisation des logiciels open sources de la gestion de bibliothèque va contribuer efficacement dans la bibliothéconomie universitaire ;
- L'informatisation du Service d'Information et de Documentation de l'Université redynamise le service et suscite de nouvelles compétences.
- L'adoption d'un ERP open source va contribuer efficacement dans le management de toutes les transactions de l'université ;
- L'utilisation de référentiel de connaissances va augmenter le rendement universitaire ;
- L'adoption des outils d'analyse de données massives va permettre de mieux gérer des données volumineuses de l'université.

Nos perspectives à court terme concernent la mise en place d'un système qui permettra d'analyser les données massives provenant des différentes sources (structurées, semi et non structurées) d'un tel établissement en vue de finaliser notre application en utilisant Hadoop. À long termes, les perspectives visent de disposer d'un système ou d'autres modules seront développés à même de mettre à la disposition des décideurs de l'université un ensemble de données pertinentes et précises pour une prise de décision facile et efficace.

Bibliographie

- [1] R. Likert, "A Technique for the Measurement of Attitudes", *Archives of Psychology*, vol. 140, p. 1–55, 1932.
- [2] A. Piekara, S. Dziuba et B. Kopeć, "The use of Ishikawa diagram as means of improving the quality of hydraulic nipple", *chapter 15. [In:] Toyotarity. Quality and machines operating conditions, Monograph. Ed. S. Borkowski, J. Selejd*, pp. 162-175, 2012.
- [3] G. Balmissé, "Gestion des connaissances: outils et applications du knowledge management", *collection entreprendre informatique*, p. 266, 2002.
- [4] R. Reix, "Systèmes d'Information et Management des Organisations", *Vuibert*, vol. 5, p. 487, 2004.
- [5] K. Laudon et L. Laudon, "Management des systèmes d'information", *Pearson Education France*, 2010.
- [6] S. Tannenbaum, "Human Resource Information Systems: User Group Implications", *Journal of Systems Management*, vol. 41, n° 11, p. 27, 1990.
- [7] P. P. S. Chen, "The entity-relationship model—toward a unified view of data", *ACM Transactions on Database Systems (TODS)*, vol. 1, n° 11, pp. 9-36, 1976.
- [8] E. Codd, S. Codd et C. Salley, "Providing OLAP (On-Line Analytical Processing) to User Analyst: An IT Mandate", vol. 32, 1993.
- [9] W. H. Inmon et R. D. Hackathorn, "Using the data warehouse", *Prism Tech Topic*, 1994.
- [10] C. Duby et S. Robin, "Analyse en composantes principales", *Institut National Agronomique, Paris-Grignon*, 2006.
- [11] Y. Renisio et R. Sinthon, "L'analyse des correspondances multiples au service de l'enquête de terrain", *Genèses*, vol. 4, pp. 109-125, 2014.
- [12] J. P. Benzécri, "L'analyse des données", *Paris: Dunod*, vol. 2, p. 1, 1973.
- [13] B. Escofie et J. Pagès, "Analyse factorielles simples et multiples : Objectifs, méthodes et interprétation", *Broché*, 1998.
- [14] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations", *Proceedings of the Fifth Symposium on Math, Statistics and Probability*, pp. 281-297, 1967.
- [15] L. Lebart, A. Morineau et M. Piron, "Statistique exploratoire multidimensionnelle", *Paris: Dunod*, vol. 3, 1995.
- [16] M. P. Gawande et D. G. Agrawal, "Face recognition using PCA and different distance classifiers", *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)*, vol. 9, n° 11, pp. 1-5, 2014.
- [17] B. Liu, "Web data mining: Exploring hyperlinks", *New York: Springer*, 2011.
- [18] M. Tsytsarau et T. Palpanas, "Survey on mining subjective data on the web", *Data Mining and Knowledge Discovery*, n° 124, p. 478–514, 2012.
- [19] E. Mathieu-Dupas, "Algorithme des k plus proches voisins pondérés et application en diagnostic", *In 42èmes Journées de Statistique*, 2010.
- [20] D. Bessalov, B. Bai, Y. Qi et A. Shokoufandeh, "Sentiment classification based on supervised latent n-gram analysis", *In Proceedings of the 20th ACM international conference on information and knowledge management*, pp. 375-382, 2011.

- [21] L.-S. Chen, C.-H. Liu et H.-J. Chiu, "A neural network based approach for sentiment classification in the blogosphere", *Journal of Informetrics*, n° 15, p. 313–322, 2011.
- [22] W. Claster, D. Hung et S. Shanmuganathan, "Unsupervised artificial neural nets for modeling movie sentiment", *In International conference on computational intelligence, communication systems and networks*, pp. 349-354, 2010.
- [23] B. Pang et L. Lee, "Opinion mining and sentiment analysis", *Foundations and Trends in Information Retrieval*, n° 12, p. 1–135, 2008.
- [24] M. Tsytsarau et T. Palpanas, "Survey on mining subjective data on the web", *Data Mining and Knowledge Discovery*, n° 124, p. 478–514, 2012.
- [25] N. B. Amor, S. Benferhat et Z. Elouedi, "Réseaux bayésiens naïfs et arbres de décision dans les systèmes de détection d'intrusions", *Technique et Science Informatiques*, vol. 25, n° 12, pp. 167-196, 2006.
- [26] J. Dinet et J. Rouet, "La recherche d'information : processus cognitifs, facteurs de difficultés et dimension de l'expertise, Interaction homme-machine et recherche d'information", pp. 133–161, 2002.
- [27] S. Dennis, P. Bruza et R. McArthur, "Web searching : A process-oriented experimental study of three interactive search paradigms", *Journal of the American*, n° 153, p. 120–133, 2002.
- [28] G. Kembellec, "Bibliographies scientifiques : de la recherche d'informations à la production de documents normés", *Sciences de l'information et de la communication*, Université Paris VIII Vincennes-Saint Denis, 2012.
- [29] L. Tanguy, "Traitement Automatique de la Langue Naturelle et interprétation : Contribution à l'élaboration d'un modèle informatique de la Sémantique Interprétative", *Informatique et langage*, Université de Rennes 1, 1997.
- [30] M. F. Porter, "An Algorithm for Suffix Stripping", *Program*, vol. 14, n° 13, p. 130–137, 1980.
- [31] X. Zhou, X. Tao, J. Yong et Z. Yang, "Sentiment analysis on tweets for social events", *In Computer Supported Cooperative Work in Design (CSCWD), 2013 IEEE 17th International Conference on, IEEE*, pp. 557-562, 2013.
- [32] A. Tumasjan, T. O. Sprenger, P. Sandner et I. Welpe, "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment", *ICWSM*, vol. 10, pp. 178-185, 2010.
- [33] S. Asur et B. Huberman, "Predicting the future with social media", *In Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on, IEEE*, vol. 1, pp. 492-499, 2010.
- [34] T. K. Das, D. Acharjya et M. R. Patra, "Opinion mining about a product by analyzing public tweets in Twitter", *In Computer Communication and Informatics (ICCCI), 2014 International Conference on, IEEE*, pp. 1-4, 2014.
- [35] M. M. Mostafa, "More than words: Social networks' text mining for consumer brand sentiments", *Expert Systems with Applications*, vol. 40, n° 110, pp. 4241-4251, 2013.
- [36] S. Mohammad, "From once upon a time to happily ever after: Tracking emotions in mail and books", *Decision Support Systems*, vol. 53, p. 730–741, 2012.
- [37] W. Medhat, A. Hassan et H. Korashy, "Sentiment analysis algorithms and applications: A survey", *Ain Shams Engineering Journal*, vol. 5, n° 14, pp. 1093-1113, 2014.
- [38] O. Kaser et D. Lemire, "Tag-cloud drawing: Algorithms for cloud visualization", *In WWW '07 Workshop on Tagging and Metadata for Social Information Organization*, 2007.

- [39] C. Seifert, B. Kump, W. Kienreich, G. Granitzer et M. Granitzer, "On the beauty and usability of tag clouds", *In 12th International Conference on Information Visualisation, IV '08*, p. 17–25, 2008.
- [40] K. Koh, B. Lee, B. Kim et J. Seo, "Maniwordle: Providing flexible control over wordle", *IEEE Transaction on Visualization and Computer Graphics*, vol. 16, n° 16, pp. 1190-1197, 2010.
- [41] H. Strobel, M. Spicker, A. Stoffel, D. Keim et O. Deussen, "Rolledout Wordles: A heuristic method for overlap removal of 2D data representatives", *Computer Graphics Forum*, vol. 31, n° 13, p. 1135–1144, 2012.
- [42] M. Baralt, S. Pennestri et M. Selvandin, "Action research: Using wordles to teach foreign language writing", *Language Learning & Technology*, vol. 15, n° 2, p. 12–22, 2011.
- [43] Y. Hassan-Montero et V. Herrero-Solana, "Improving tag-clouds as visual information retrieval interfaces", *In International Conference on Multidisciplinary Information Sciences and Technologies, InSciT '06*, p. 25–28, 2006.
- [44] V. Gupta et G. S. Lehal, "A survey of text mining techniques and applications", *Journal of emerging technologies in web intelligence*, vol. 1, n° 11, pp. 60-67, 2009.
- [45] Y.-X. Chen, R. Santamaría, A. Butz et R. Thérion, "TagClusters: Semantic aggregation of collaborative tags beyond TagClouds", *In 10th International Symposium on Smart Graphics, SG '09, Springer*, pp. 56–67, 2009.
- [46] F. V. Paulovich, F. M. B. Toledo, G. P. Telles, R. Minghim et L. G. Nonato, "Semantic wordification of document collections", *Computer Graphics Forum*, vol. 31, n° 13, pp. 1145–1153, 2012.
- [47] Y. Wu, T. Provan, F. Wei, S. Liu et K.-L. Ma, "Semantic-preserving word clouds by seam carving", *Computer Graphics Forum*, vol. 30, n° 13, pp. 741–750, 2011.
- [48] K. Fujimura, S. Fujimura, T. Matsubayashi, T. Yamada et H. Okuda, "Topigraphy: visualization for large-scale tag clouds", *In International Conference on World Wide Web, WWW '08*, p. 1087–1088, 2008.
- [49] M. Burch, S. Lohmann, D. Pompe et D. Weiskopf, "Prefix tag clouds", *In 17th International Conference on Information Visualisation, IV '13, IEEE*, p. 45–50, 2013.
- [50] F. Heimerl, S. Lohmann, S. Lange et T. Ertl, "Word cloud explorer: Text analytics based on word clouds", *In 47th Hawaii International Conference on System Sciences, HICSS '14, IEEE*, p. 1833–1842, 2014.
- [51] S. Lohmann, F. Heimerl, F. Bopp, M. Burch et T. Ertl, "Concentri cloud: Word cloud visualization for multiple text documents", *Information Visualisation (iV), IEEE*, p. 114–120, 2015.
- [52] Q. Castellà et C. Sutton, "Word storms: Multiples of word clouds for visual comparison of documents", *In 23rd International Conference on World Wide Web, WWW '14, ACM*, p. 665–676, 2014.
- [53] C. Collins, M. S. T. Carpendale et G. P. Docuburst, "Visualizing document content using language structure", *Computer Graphics Forum*, vol. 28, n° 13, p. 1039–1046, 2009.
- [54] M. Burch, S. Lohmann, F. Beck, N. Rodriguez, L. D. Silvestro et D. Weiskopf, "RadCloud: Visualizing multiple texts with merged word clouds", *In 18th International Conference on Information Visualisation, IV '14, IEEE*, p. 108–113, 2014.
- [55] P. J. Sadalage et M. Fowler, "NoSQL distilled: a brief guide to the emerging world of polyglot persistence", 2013.

- [56] J. Han, E. Haihong, G. Le et J. Du, "Survey on NoSQL database", chez *In 2011 6th international conference on pervasive computing and applications, IEEE*, 2011.
- [57] J. Campos, "A big data analytical architecture for the Asset Management", *Procedia CIRP 64*, p. 369 – 374, 2017 .
- [58] G. Lakhshen, S. Vranes et V. Janev, big data and quality : a littérature review, *2016 24th Telecommun Forum*, pp. 1-4, 2016.
- [59] E. McNulty, "Understanding Big data :The seven V's", *Dataconomy*, 2014.
- [60] R. Chinchilla-Arley et M. Fernández Morales, M. "Bibliotecas automatizadas con software libre: establecimiento de niveles de automatización y clasificación de las aplicaciones", *Bibliotecas*, vol. 30, n° 2, 2012.
- [61] G. Biswas et D. Paul, "An evaluative study on the open source digital library softwares for institutional repository", *International Journal of Library and Information Science*, vol. 2, n° 11, pp. 1-10, 2009.
- [62] T. T. MULLER, "Étude comparative des principaux SIGB", p. 23, 2008.
- [63] V. Mayer-Schönberger et K. Cukier, "Big Data: La révolution des données est en marche", *Robert Laffont*, 2014.
- [64] E. Devarissias, "La gestion de l'informatisation des périodiques papiers de la bibliothèque universitaire de l'université de Lille 1", *Projet professionnel.-Lyon: ENSSIB*, 2005.
- [65] P. Gottschalk, "Strategic knowledge management technology", *Macmillan*, 2005.
- [66] I. Feinerer et K. Hornikx, "tm: Text Mining Package", *R package version 0.7-3*, 2017.
- [67] M. Bouchet-Valat, "SnowballC: Snowball stemmers based on the C libstemmer UTF-8 library", *R package version 0.5.1*, 2014.
- [68] I. Fellows, "wordcloud: Word Clouds", *R package version 2.5*, 2014.
- [69] E. Neuwirth, "RColorBrewer: ColorBrewer Palettes", *R package version 1*, pp. 1-2, 2014.
- [70] A. Hamouda et M. Rohaim, "Reviews classification using sentiwordnet lexicon", *In World congress on computer science and information technology*, vol. 23, pp. 104-105, 2011.
- [71] H. Wickham, "ggplot2: Elegant Graphics for Data Analysis", *Springer-Verlag New York*, 2009.
- [72] G. Golemund et H. Wickham, "Dates and Times Made Easy with lubridate", *Journal of Statistical Software*, vol. 40, n° 13, pp. 1-25, 2011.
- [73] N. Giallousis, "Scale: Likert Type Questionnaire Item Analysis", *R package version 1.0.4*, 2015.
- [74] H. Wickham, "stringr: Simple, Consistent Wrappers for Common String Operations", *R package version 1.2.0*, 2017.
- [75] M. Jockers, "Syuzhet: Extract Sentiment and Plot Arcs from Text", 2015.

Glossaire

Terme	Signification
APOGEE	Application pour l'organisation et la gestion des enseignements et des étudiants (inscription des étudiants, suivi des parcours, notes ...)
Bi data	Données massives ou données volumineuses
Bibliothèque traditionnelle	bibliothèque qui adopte le système de prêt manuelle sans utiliser le système informatique.
Camambert	Diagramme circulaire
Lemmatisation	l'analyse d'un texte en se basant sur le lexique dans le but de regrouper les mots d'une même famille
MOA	Maitrise d'Ouvrage est l'entité porteuse du besoin, celui qui définit les objectifs, le calendrier et le budget du projet
MOE	Maitrise d'œuvre une personne physique qui est responsable de la bonne conduite du projet
Moodle	Plateforme d'apprentissage en ligne
NoSQL	Une nouvelle nomination des données actuelles combinent des données structurés et non structurés.
OBS	Structure organisationnelle du projet est un schéma représente les responsabilités de chaque membre
Protocole Z39.50	c'est un protocole de communication informatique utilisé en réseau informatique notamment pour interroger des catalogues des bibliothèques numériques
RACI	C'est une matrice regroupant l'WBS (taches) et l'OBS (acteurs) en affectant à chaque acteur du projet 1 ou plusieurs lettres parmi les lettres R, A, C ou I avec R est le Responsable, A est l'Acteur, C qui doit être consulté et I qui doit être informé.
SMART	Un objectif SMART est un objectif Simple, Mesurable, Ambition, Réaliste et définie dans le Temps
Sphinx	Un logiciel d'enquêtes et d'analyse statistique de données
SPSS	(Abréviation de Statistical Package for the Social Sciences) est un logiciel pour l'analyse statistique de données statistiques
SWOT	C'est un outil d'analyse stratégique, il combine un diagnostic interne de l'entreprise (Forces S et Faiblesses W) et un diagnostic externe (Opportunités O, et les Menaces T)
Visual Timetabling	Application qui permet de gérer des emplois du temps
WBS	organigramme des tâches du projet est une décomposition hiérarchique des travaux nécessaires pour réaliser les objectifs du projet
XLSTAT	C'est un logiciel d'analyse de données statistique pour Microsoft

Annexe

Annexe 1: Quelques SGBD

Tableau 23 : Liste des SGBD existe sur le marché

Progiciel	Éditeur	Licence	Site web
Oracle	Oracle corporation	Propriétaire	http://www.oracle.com/
Microsoft SQL Server	Microsoft	Propriétaire	https://www.microsoft.com/fr-fr/sql-server/default.aspx
Informix	IBM	Propriétaire	http://www-01.ibm.com/software/data/informix/
MYSQL	MYSQL AB	GPL	http://www.MySQL.com/
PostgreSQL	Michael Stonebraker	BSD	http://www.postgresql.org/
Microsoft Access	Microsoft	Propriétaire	https://products.office.com/fr-fr/access
SQLite	D. Richard Hipp	Domain public	https://www.sqlite.org/
IBM DB2	IBM	Propriétaire	http://www-01.ibm.com/software/data/db2/

Annexe 2 : TF-IDF

TF : c'est la fréquence des termes dans un document (Term frequency en anglais)

IDF : Fréquence Inverse du document(en anglais Inverse Document Frequency) : utiliser pour mesurer l'importance d'un terme dans l'ensemble du corpus de documents.

$$TF_{ij} = \frac{n_{ij}}{N} \quad (6)$$

Avec n_{ij} est le nombre total du terme i dans le document j et N est le nombre total des mots.

$$IDF_{ij} = \log \frac{D}{D_t} \quad (7)$$

Avec D est le nombre total des documents.

D_t est le nombre de document ou il apparait le terme t .

$$TFIDF_{ij} = TF_{ij} \times IDF_{ij} \quad (8)$$

Annexe 3 : Niveau de satisfaction

Tableau 24: Niveau de satisfaction concernant les heures d'ouverture, la réception du personnel et l'état des locaux de la bibliothèque

Niveau de satisfaction	Faculté	Satisfaction	Pourcentage %
Pas satisfait	FLSH	l'horaire d'ouverture de la bibliothèque	0
		La réception du staff	11
		État des locaux de la bibliothèque	5
	FST	l'horaire d'ouverture de la bibliothèque	36
		La réception du staff	11
		État des locaux de la bibliothèque	26
	FP	l'horaire d'ouverture de la bibliothèque	64
		La réception du staff	78
		État des locaux de la bibliothèque	68
Moyennement satisfait	FLSH	l'horaire d'ouverture de la bibliothèque	8
		La réception du staff	10
		État des locaux de la bibliothèque	5
	FST	l'horaire d'ouverture de la bibliothèque	12
		La réception du staff	50
		État des locaux de la bibliothèque	35
	FP	l'horaire d'ouverture de la bibliothèque	80
		La réception du staff	40
		État des locaux de la bibliothèque	59
Satisfait	FLSH	l'horaire d'ouverture de la bibliothèque	3
		La réception du staff	0
		État des locaux de la bibliothèque	0
	FST	l'horaire d'ouverture de la bibliothèque	38
		La réception du staff	39
		État des locaux de la bibliothèque	24
	FP	l'horaire d'ouverture de la bibliothèque	59
		La réception du staff	61
		État des locaux de la bibliothèque	76

Niveau de satisfaction	Faculté	Satisfaction	Pourcentage %
Très satisfaisant	FLSH	l'horaire d'ouverture de la bibliothèque	7
		La réception du staff	5
		État des locaux de la bibliothèque	0
	FST	l'horaire d'ouverture de la bibliothèque	27
		La réception du staff	15
		État des locaux de la bibliothèque	0
	FP	l'horaire d'ouverture de la bibliothèque	64
		La réception du staff	79
		État des locaux de la bibliothèque	100

Annexe 4 : Configuration d'EasyPHP

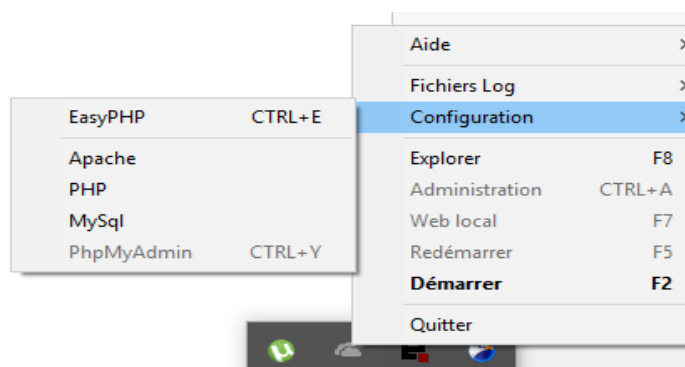


Figure 78 Menu déroulant d'EasyPhp

Pour arriver aux fichiers de configuration, faire clic droit de la souris sur l'icône de EasyPhp, un menu déroulant apparaît. Choisir l'option Configuration et dans le sous-menu, choisir Php pour modifier le fichier **php.ini**.

❖ Paramètres à modifier dans le fichier **php.ini**

- `max_execution_time = 300;`
- `max_input_time = 600;`
- `memory_limit = 128 M;`
- `post_max_size = 128 M;`

- `upload_max_filesize = 128 M`
- ❖ Paramètre à modifier dans **my.ini**
- `max_allowed_packet = 128 M ;`

Dans le fichier `php.ini`, il faut enlever le point-virgule (;) devant les extensions suivantes pour qu'on puisse activer les paramètres de PHP.

- `extension=php_bz2.dll ; # sauvegardes de la base`
- `extension=php_curl.dll ;`
- `extension=php_gd2.dll ; # génération de code-barres`
- `extension=php_mbstring.dll ;`
- `extension=php_mime_magic.dll ; # permet de récupérer les types de fichier des documents numériques de façon propre`
- `extension=php_MySQL.dll;`
- `extension=php_sockets.dll;`
- `extension=php_xsl.dll ; # Pour faire la sauvegardes de la base`
- `extension=php_zip.dl`
- `extension= php_yaz.dll ; # recherche z3950`

Annexe 5 : Critère de choix de la version de PMB et paramètres

- Php 5.2 – MySQL 5.1 pour PMB $\leq 3.5.1$ (ex. EasyPhp 5.2.10)
- Php 5.3 – MySQL 5.1 pour PMB $\geq 3.5.1$ (ex. EasyPhp 12.1, Wamp 2.2.d ou Xamp 1.7.7)
- Donc, pour PMB 3.5.1, vous avez le choix entre le Php 5.2 et le Php 5.3
- Php 5.3 et MySQL 5.1 deviennent obligatoires pour PMB ≥ 4.0 .
- Php 5.4 est possible seulement à partir de la version 4.0.8 de PMB
- Php 5.4 (ou 5.5 ou 5.6) est obligatoire à partir de la version 4.2 de PMB

Paramètres système

Nous avons besoin des informations de connexion au serveur en tant qu'administrateur afin de réaliser toutes les opérations de création de la base de données :

Utilisateur MySQL :
Mot de passe :
Serveur :
Base de données :

Si vous remplissez "Base de données", la rubrique "Paramètres PMB" ci-dessous sera ignorée : les tables de PMB seront créées dans la base de données renseignée, par exemple de votre hébergement.

Paramètres PMB

Si vous n'avez pas précisé de base de données à la rubrique précédente, vous devez préciser ici l'utilisateur MySQL et son mot de passe qui seront utilisés par PMB pour se connecter à la base dont le nom doit être renseigné également.

Utilisateur PMB : **Fixé par les paramètres système**
Mot de passe : **Fixé par les paramètres système**
Base de données PMB : **Fixé par les paramètres système**

Attention si une base portant le même nom existe déjà, elle sera détruite, et les tables qu'elle contient définitivement perdues.

Figure 79 Paramétrage du système et de la base de données PMB

Annexe 6 : Fiche d'importation des exemplaires

	ISBN	CB	COTE	TITR1	AUTEUR	editeur nom	editeur Ville	Date parution	support	Nombre page	prix	langue	materiel accompagnement
1													
2													
3													

Figure 80 Modèle de création des notices bibliographiques en format .csv

Conversions/Export > Conversions de fichiers externes

Conversions de fichiers externes Export de la base PMB Paramètres d'export Gestion Paramètres d'export OPAC

Conversion Text -> UNIMARC ISO2709 terminée.

100%
1 notices traitées

Destination du fichier

Lancer l'import des notices à partir de ce fichier
 Lancer l'import des exemplaires à partir de ce fichier
 Télécharger le fichier converti

Pour l'import, choisir la fonction à utiliser: Spécifique aux bases bretagne et échanges mémo-notice

Terminer l'étape de conversion

Figure 81 Conversion du fichier exemplaires en format Unimarc ISO2709

Annexe 7 : Enquête de satisfaction des étudiants

Bonjour, nous menons un projet de recherche universitaire afin de mieux connaître le niveau de satisfaction des services présentés par la bibliothèque. Merci de nous aider dans notre démarche en répondant aux questions suivantes :

Q1. Vous êtes :

- Enseignants-chercheurs
- 3^{ème} cycle
- Cycle master
- Cycle Licence
- Administrateur

Q2. Votre Établissement ?

- FST
- FP
- FLSH

Q3. Sexe ?

- M
- F

Q4. Quel est votre tranche d'âge ?

- < 20
- [20-24]
- [25 -29]
- [30-34]
- [35-39]
- [40 -44]
- >= 45

Q5. Boursier ou non ?

- Oui
- Non

Q6. Domaine d'étude ?

- Physique
- Informatique
- Mathématique
- Arabe
- Française
- Anglaise

- Géographie
- Biologie
- Géologie
- Chimie
- Islamique
- Économie
- Histoire
- Droit
- Autres

Q7. Quel est votre fréquence des visites à la bibliothèque ?

- Très souvent
- Parfois
- Rarement
- Jamais

Q8. Quel est le niveau de satisfaction des services offerts par la bibliothèque traditionnelle ?

- Très satisfaisant
- Satisfaisant
- Moyennement satisfaisant
- Pas satisfaisant du tout

Q9. Quels sont les types des documents les plus utilisés ?

- Livres
- Thèses et mémoires
- Périodiques
- Articles
- Autres

Q10. Est-ce que vous trouvez toujours le document demandé dans le fond documentaire ?

- Oui
- Non
- Rarement

Q11. Avez-vous déjà visité une bibliothèque informatisée ?

- Oui
- Non

Q12. Quel est votre avis sur le projet d'informatisation de la bibliothèque ?

- Important
- Pas nécessaire

Q13. Quelles fonctionnalités souhaitées vous informatiser le plus vite ?

- Gestion de prêts
- Gestion des lecteurs
- Acquisition(Commandes)
- Bulletinage (abonnements)
- Catalogue des monographies (mémoires et thèses)
- Partage entre bibliothèque de l'USMS
- Statistiques du fond documentaire
- Autres

Q14. Combien de temps vous passez pour emprunter un livre ?

- <5min
- [5min ; 10min[
- [10min ; 15min[
- [15min ; 20min[
- >=20min

Q15. Combien de temps vous passez à la bibliothèque par jour ?

- <1h
- 1h
- 2h
- 3h
- 4h
- 5h
- 6h
- 7h
- 8h

Q16. Motif de visite la bibliothèque ?

- Réviser les cours
- Consultation sur place les livres sans les empruntés
- Emprunter les livres
- Profiter d'un moment libre
- Allez sur internet
- Retrouver des amis
- Visionner un film
- Taper un texte

- Consultation des ouvrages en ligne de la bibliothèque de l'USMS
- Autres
- Q17. Comment vous pouvez chercher un document ?
- Consulte le catalogue via le site de la biblio de l'USMS
- Aidez par un bibliothécaire
- Aidez par un ami
- Aidez par le professeur
- Q18. Vous empruntez des livres conseillés par ton professeur ?
- Oui
- Non
- Rarement
- Q19. Vous achetez des livres ?
- Oui
- Non
- Rarement
- Q20. Vous lisez des documents d'autres disciplines ?
- Oui
- Non
- Rarement
- Q21. Vous avez un ordinateur ?
- Oui
- Non
- Q22. Quel moyen avec lequel vous utilisez pour l'accès à l'internet ?
- Via l'internet mobile
- Via la connexion de la Faculté
- Via la connexion installée à la maison
- Je n'utilise pas internet
- Q23. Vous utilisez des moyens informatiques pour communiquer avec votre professeur ?
- Oui
- Non
- Rarement
- Q24. Vous utilisez la documentation électronique?
- Oui
- Non
- Rarement

Q25. Vous maîtrisez les bases de données ?

- Oui
- Non

Q26. Vous avez suivi une formation de la recherche documentaire ?

- Oui
- Non

Q27. Comment vous trouvez l'état des locaux de la bibliothèque ?

- Très satisfaisant
- Satisfaisant
- Moyennement satisfaisant
- Pas satisfaisant du tout

Q28. Comment vous trouvez l'horaire d'ouverture de la bibliothèque ?

- Très satisfaisant
- Satisfaisant
- Moyennement satisfaisant
- Pas satisfaisant du tout

Q29. Comment vous trouvez la réception du staff de la bibliothèque ?

- Très satisfaisant
- Satisfaisant
- Moyennement satisfaisant
- Pas satisfaisant du tout

Q30. Votre établissement organise des activités au sein de la bibliothèque ?

- Oui
- Non
- Rarement

Q31. Donnez tout autre commentaire ou suggestion que vous désirez ?