



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE
LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ ABDELHAMID IBN BADIS - MOSTAGANEM

Faculté des Sciences Exactes et de l'Informatique
Département de Mathématiques et d'Informatique
Filière : Informatique

MEMOIRE DE FIN D'ETUDES
Pour l'Obtention du Diplôme de Master en Informatique
Option : **Ingénierie des systèmes d'information**

THEME :

**Amélioration de l'évaluation d'un outil de calcul de la qualité de
référencement d'articles scientifiques**

Etudiant(e) : M^{elle} « BENANI Imene »

Encadrant(e): Dr «FILALI Fatima Zohra»

Année Universitaire 2018-2019

Résumé

Toute activité humaine est évaluée. Le comportement d'un individu ou d'un groupe est jugé. Le travail de chacun est estimé par la hiérarchie ou par les collègues de travail. Une entreprise est appréciée par ses clients qui achètent ses produits ou ses services. Une administration est bien ou mal considérée par les citoyens. Les hommes politiques sont jugés par les électeurs. Quant à l'activité scientifique, elle est évaluée ou expertisée par différents groupes : la communauté scientifique elle-même, la société, les responsables économiques et politiques

Une nouvelle connaissance, une découverte scientifique, sont avérées à partir du moment où elles sont validées par la communauté scientifique. Cette validation se fait, généralement, via une publication scientifique ou une prise de brevet. La qualité scientifique d'un chercheur, d'un laboratoire, d'un organisme de recherche ou d'une partie de la communauté scientifique est donc appréciée via son évaluation scientifique.

Les observations et les découvertes des scientifiques, interprétées dans le cadre de modèles de pensée, font l'objet de publications écrites ou orales. Pour la recherche académique, la forme la plus répandue de publication est l'article où sont consignées les expériences, observations et interprétations des auteurs.

Ce Mémoire de fin d'études présente l'évaluation de la recherche scientifique ainsi que les techniques et les stratégies d'évaluation. Ce travail cite aussi les facteurs qu'on peut utiliser pour connaître la qualité des articles scientifiques les deux derniers chapitres présente notre solution proposé pour la qualité de référencement des articles scientifiques. L'objectif de ce projet est d'améliorer l'évaluation d'un outil de calcul de la qualité de référencement des articles scientifiques, c'est ce que nous verrons dans les sections suivantes.

Les mots-clés : Recherche scientifiques, bibliométrie, Evaluation de la Recherche, qualité, facteurs, articles, indicateurs.

Abstract:

All human activity is evaluated. The behavior of an individual or group is judged. The work of each is estimated by the hierarchy or work colleagues. A company is appreciated by its customers who buy its products or services. An administration is well or badly considered by citizens. Politicians are judged by voters. As for scientific activity, it is evaluated or appraised by different groups: the scientific community itself, society, economic and political leaders.

A new knowledge, a scientific discovery, are proven from the moment they are validated by the scientific community. This validation is usually done via a scientific publication or a patent. The scientific quality of a researcher, a laboratory, a research organization or a part of the scientific community is therefore appreciated through its scientific evaluation.

The observations and discoveries of scientists, interpreted in the framework of models of thought, are the subject of written or oral publications. For academic research, the most

popular form of publication is the article in which authors' experiences, observations and interpretations are recorded.

This dissertation presents the evaluation of scientific research as well as evaluation techniques and strategies. This work also cites the factors that can be used to know the quality of the scientific articles. The last two chapters present our proposed solution for the referencing quality of scientific articles. The objective of this project is to improve the evaluation of a tool for calculating the quality of referencing scientific articles, which we will see in the following sections.

Keys-words: Scientific research, bibliometric, evaluation of research, quality, factors, articles, indicators.

DEDICACE

*Je dédie ce **modeste travail***

A mes parents. Aucun hommage ne pourrait être à la hauteur de l'amour Dont ils ne cessent de me combler. Que dieu leur procure bonne santé et longue vie.

A l'homme de ma vie, mon exemple éternel, mon soutien moral et source de joie et de bonheur, celui qui s'est toujours sacrifié pour me voir réussir, que dieu te garde dans son vaste paradis, à toi mon père.

*A la lumière de mes jours, la source de mes efforts, la flamme de mon cœur, ma vie et mon bonheur, **maman** que j'adore.*

Aux personnes dont j'ai bien aimé la présence dans ce jour, à tous mes sœurs, à mes nièces et mes neveux , je dédie ce travail dont le grand plaisir leurs revient en premier lieu pour leurs conseils, aides, et encouragements.

*Je tiens à exprimer toute ma reconnaissance à mon **Directeur de mémoire**, Madame **FILALI Fatima Zohra** je la remercie de m'avoir encadré, orienté, aidé et conseillé. A tous mes professeurs.*

Aux personnes qui m'ont toujours aidé et encouragé, qui étaient toujours à mes côtés, et qui m'ont accompagnaient durant mon chemin d'études supérieures, à tous mes amis, mes collègues d'étude.

Merci 

REMERCIEMENT

Tout d'abord, je tiens à remercier le bon **Dieu** le tout **Puissant** de m'avoir donné la force et le courage de mener à bien ce travail, également je remercie infiniment **mes parents**, qui m'ont encouragé et aidé à arriver à ce stade de ma formation

Ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadrement de **Dr FILALI Fatima Zohra**, je la remercie pour la qualité de son encadrement exceptionnel, pour sa patience, sa rigueur et sa disponibilité durant ma préparation de ce mémoire.

Je remercie Mon fiancé **Bachir Maïza** qui m'a encouragé et pour son soutien quotidien, et **Benchehida Chawki** pour son aide, ses conseils et ses réponses.

Je tiens à remercier tous ceux et celle qui ont contribué à finaliser ce travail. Mes profonds remerciements vont également à toutes les personnes qui m'ont aidé et soutenue de près ou de loin à la réalisation de ce travail.

Merci 

Résumé

Toute activité humaine est évaluée. Le comportement d'un individu ou d'un groupe est jugé. Le travail de chacun est estimé par la hiérarchie ou par les collègues de travail. Une entreprise est appréciée par ses clients qui achètent ses produits ou ses services. Une administration est bien ou mal considérée par les citoyens. Les hommes politiques sont jugés par les électeurs. Quant à l'activité scientifique, elle est évaluée ou expertisée par différents groupes : la communauté scientifique elle-même, la société, les responsables économiques et politiques

Une nouvelle connaissance, une découverte scientifique, sont avérées à partir du moment où elles sont validées par la communauté scientifique. Cette validation se fait généralement, via une publication scientifique ou une prise de brevet. La qualité scientifique d'un chercheur, d'un laboratoire, d'un organisme de recherche ou d'une partie de la communauté scientifique est donc appréciée via son évaluation scientifique.

Les observations et les découvertes des scientifiques, interprétées dans le cadre de modèles de pensée, font l'objet de publications écrites ou orales. Pour la recherche académique, la forme la plus répandue de publication est l'article où sont consignées les expériences, observations et interprétations des auteurs.

Ce Mémoire de fin d'études consiste à étudier l'évaluation de la recherche scientifique ainsi que les techniques et les stratégies d'évaluation. Cette étude porte sur les facteurs utilisés pour reconnaître la qualité des articles scientifiques. A travers cette étude, nous essayons de présenter une solution pour évaluer la qualité de référencement des articles scientifiques.

Les mots-clés : Recherche scientifiques, bibliométrie, Valuation de la Recherche, qualité, facteurs bibliométriques, articles, indicateurs, bases de données.

Abstract

All human activity is evaluated. The behavior of an individual or group is judged. The work of each is estimated by the hierarchy or work colleagues. A company is appreciated by its customers who buy its products or services. An administration is well or badly considered by citizens. Politicians are judged by voters. As for scientific activity, it is evaluated or appraised by different groups: the scientific community itself, society, economic and political leaders.

A new knowledge, a scientific discovery, are proven from the moment they are validated by the scientific community. This validation is usually done via a scientific publication or a

patent. The scientific quality of a researcher, a laboratory, a research organization or a part of the scientific community is therefore appreciated through its scientific evaluation.

The observations and discoveries of scientists, interpreted in the framework of models of thought, are the subject of written or oral publications. For academic research, the most popular form of publication is the article in which authors' experiences, observations and interpretations are recorded.

This dissertation presents the evaluation of scientific research as well as evaluation techniques and strategies. This study cites the factors that can be used to know the quality of the scientific articles. Through this study, we present a solution for the evaluating the quality of scientific articles.

Keys-words: Scientific research, bibliometric, evaluation of research, quality, factors, articles, indicators.

DEDICACE

Je dédie ce modeste travail

A mes parents. Aucun hommage ne pourrait être à la hauteur de l'amour dont ils ne cessent de me combler. Que dieu leur procure bonne santé et longue vie.

A l'homme de ma vie, mon exemple éternel, mon soutien moral et source de joie et de bonheur, celui qui s'est toujours sacrifié pour me voir réussir, que dieu te garde dans son vaste paradis, à toi mon père.

A la lumière de mes jours, la source de mes efforts, la flamme de mon cœur, ma vie et mon bonheur, maman que j'adore.

Aux personnes dont j'ai bien aimé la présence dans ce jour, à toutes mes sœurs, à mes nièces et mes neveux, je dédie ce travail dont le grand plaisir leurs revient en premier lieu pour leurs conseils, aides, et encouragements.

*A mon fiancé pour sa confiance, amour et sacrifice **Maiza Bachir**.*

*Je tiens à exprimer toute ma reconnaissance à mon Directeur de mémoire, Madame **FILALI Fatima Zohra**, je la remercie de m'avoir encadrée, orientée, aidée et conseillée. A tous mes professeurs.*

Aux personnes qui m'ont toujours aidée et encouragée, qui étaient toujours à mes côtés, et qui m'ont accompagnée durant mon chemin d'études supérieures, à tous mes amis et mes collègues d'étude.

Merci 

REMERCIEMENT

Tout d'abord, je tiens à remercier le bon **Dieu** le tout **Puissant** de m'avoir donné la force et le courage de mener à bien ce travail. Je remercie infiniment **mes parents**, qui m'ont encouragée et aidée à arriver à ce stade de ma formation.

Ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadrement de **Dr FILALI Fatima Zohra**, je la remercie pour la qualité de son encadrement exceptionnel, pour sa patience, sa rigueur et sa disponibilité durant ma préparation de ce mémoire.

Je remercie mon fiancé **Bachir Maïza** pour son encouragement et son soutien quotidien, et **Benchehida Chawki** pour son aide, ses conseils et ses réponses.

Je tiens à remercier tous ceux et celles qui ont contribué à finaliser ce travail. Mes profonds remerciements vont également à toutes les personnes qui m'ont aidée et soutenue de près ou de loin à la réalisation de ce travail.

Merci 

Table des matières

Résumé.....	1
Abstract.....	1
Dédicace.....	3
Remerciement.....	4
Table des matières	5
Liste des tableaux	8
Liste des équations.....	9
Liste des Figures.....	10
Introduction générale	11
Chapitre 1 :.....	13
1.1. Introduction.....	13
1.2. Articles scientifiques	13
1.3. La recherche scientifique.....	13
1.4. Évaluation des articles scientifiques [3].....	14
1.5. Techniques d'évaluation [5]	15
1.5.1. L'observation en situation.....	15
1.5.2. L'observation participante	15
1.5.3. L'entrevue de recherche.....	15
1.5.4. Le questionnaire ou le sondage	16
1.5.5. L'expérimentation.....	17
1.5.6. L'analyse de contenu	17
1.5.7. L'analyse statistiques.....	17
1.5.8. La bibliométrie [6].....	18
1.6. Les facteurs d'évaluation de la qualité d'un article scientifique [7].....	19
1.6.1. Les facteurs de revues.....	19
1.6.2. Les facteurs pour les auteurs [8].....	20
1.6.3. Les facteurs de la publication scientifique	23
1.7. Les bases de données bibliographiques [9]	23
1.7.1. Thomson Reuters (ISI - Institute for Scientific Information)	23
1.7.2. Elsevier (Scopus).....	24

1.7.3. Google Scholar [11].....	24
1.7.4. WorldCat.....	25
1.7.5. BASE.....	25
1.7.6. Semantic Scholar.....	25
1.8. Conclusion.....	26
Chapitre 2 :.....	27
2.1. Introduction.....	27
2.2. Les techniques d'évaluation.....	27
2.2.1. L'intelligence Artificielle.....	27
2.2.2. L'apprentissage Machine [12].....	28
2.2.3. Les Algorithmes génétiques [14].....	28
2.3. Travaux existants sur l'évaluation d'articles scientifiques.....	30
2.3.1. Facteur d'impact pondéré (Weighted Impact Factor WIFE) [15].....	30
2.3.2. Médias sociaux et altimétriques [18].....	32
2.3.3. Les Algorithmes génétiques et les Réseaux gaussiens bayésiens [8] [17].....	33
2.4. Conclusion.....	36
Chapitre 3 :.....	37
3.1. Notations.....	38
3.2. Introduction.....	40
3.3. Contexte et motivation du projet.....	40
3.3.1. Problématique.....	40
3.3.2. Analyse de l'existant.....	40
3.3.3. Solution proposée.....	41
3.4. Approche de conception de la solution proposée.....	41
3.4.1. Algorithme génétique.....	42
3.4.2. Description de la solution proposée.....	42
3.4.3. Algorithmes de la solution proposée.....	43
3.4.3.1. La normalisation.....	44
3.4.3.2. La corrélation.....	44
3.4.3.3. Estimation des poids initiaux.....	47
3.4.3.4. Estimation des poids finaux.....	48
3.4.3.5. Calcul du score d'évaluation.....	49

3.5. Architecture de la solution proposée.....	49
3.6. Modélisation de la solution.....	50
3.6.1. Digramme de cas d'utilisation.....	50
3.6.2. Digramme de classe.....	51
3.6.3. Digramme d'activité.....	53
3.6.4. Digramme de séquence.....	54
3.7. Conclusion.....	55
Chapitre 4 :.....	56
4.1. Introduction.....	56
4.2. Langages de programmation.....	56
4.2.1. JavaScript.....	56
4.2.2. HTML.....	56
4.2.3. CSS.....	57
4.3. Environnement et outils de développement.....	57
4.3.1. Node.js.....	57
4.3.2. Mongo dB et Mongoose.....	57
4.4.3. Cermin.....	58
4.4. Architecture d'implémentation du système.....	58
4.5. Cas d'utilisations.....	59
4.5.1. Lancer le serveur de données mongoDB.....	59
4.5.2. Lancement du serveur de calcul.....	60
4.5.3. Lancement de l'évaluation.....	60
4.5.4. Affichage des résultats finaux.....	64
4.6. Conclusion.....	65
Conclusion général.....	66
Références bibliographiques.....	67

Liste des tableaux

Tableau 1. Quelques indicateurs du journal analysé reçu de la base de données Scopus.30

Tableau 2 : Facteurs d'impact de la citation de journaux et du nombre de leurs citations.31

Tableau 3 :Facteurs d'impact des RJC et facteurs d'impact pondérés proposés dans les journaux analysés.....31

Liste des équations

Équation 1 :	20
Équation 2	20
Équation 3 :	21
Équation 4	21
Équation 5	21
Équation 6 :	21
Équation 7 :	22
Équation 8 :	30
Équation 9	34
Équation 10	35
Équation 11	43
Équation 12	44

Liste des Figures

- Cycle d'évaluation	14
- Organigramme d'algorithme génétique.....	29
- Les étapes d'un algorithme génétique. [17] [8]	34
- Architecture générale de notre application	50
- Diagramme de cas d'utilisation.....	51
- Diagramme de classe	52
- Diagramme d'activité	54
- Diagramme de Séquence	55
- Architecture fonctionnelle	58
- Lancement de mongoDB	60
- Lancement du serveur nodejs.....	60
- Lancement de l'évaluation d'un article	62
- Traitements en cours du serveur.....	63
- Traitements en cours du serveur (suite).....	64
- Interface des résultats finaux.....	65

Introduction générale

Contexte

Le but de la recherche scientifique étant d'accroître nos connaissances et notre compréhension, c'est par le biais de la publication de nouvelles connaissances que nous pourrions compléter les connaissances globales de l'humanité. Pour la recherche financée par les pouvoirs publics, la publication des résultats doit être considérée comme une partie essentielle du processus de recherche. La recherche financée dans un but commercial peut impliquer certaines restrictions au niveau de l'édition et les droits de propriété des bailleurs de fonds peuvent être protégés par des brevets et par la protection des droits d'auteurs.

L'évaluation de ces publications est un processus fondamental, pour évaluer les projets autant que les chercheurs et les équipes de recherche, car ils jouent un rôle essentiel dans la valorisation de la recherche effectuée. Dans sa définition la plus large, une « publication scientifique » peut se faire sous n'importe quelle forme pour communiquer les résultats de la recherche au grand public (disponible gratuitement ou contre paiement), à condition qu'elle affecte un caractère permanent (ou d'archivage). Les articles dans des revues scientifiques et les livres sont traditionnellement les formes les plus communes, mais citons aussi les rapports, les pages Web et les cartes.

L'approche de l'évaluation adoptée ici se concentrera d'abord sur l'évaluation des articles scientifiques, qui constituent l'essentiel de la discussion. L'évaluation des revues utilisera tout d'abord des critères similaires, mais sera complétée par l'analyse d'un certain nombre de caractéristiques propres aux revues. Il en ira de même pour les livres et les autres publications.

Problématique

L'objectif du présent travail est d'étudier les différents facteurs et méthodes d'évaluation afin de proposer une méthode d'évaluation des références d'un article scientifique.

Organisation du mémoire

Le mémoire est composé de 4 chapitres, chacun représente une partie du travail effectué que ce soit théorique ou pratique, en voici les titres :

- Introduction générale.
- Chapitre 1 : **État de l'art** : aborde les différentes notions et concepts se rapportant à notre travail.
- Chapitre 2 : **Méthodes et stratégies d'évaluation** : présente quelques méthodes et stratégies d'évaluation d'articles scientifiques déjà existants.
Conception et développement : présente la solution proposée et explique les
- Chapitre 3 : différentes étapes de conception : algorithmes , diagrammes et architecture de l'outil proposé.
Implémentation et résultat : contient la dernière phase de notre travail qui est la
- Chapitre 4 : mise en œuvre de la solution proposée..
- Conclusion générale

Chapitre 1 :

État de l'art

1.1. Introduction

Ce chapitre constitue une introduction aux concepts et notions sur les articles scientifiques et les facteurs d'évaluations. Après un rappel sur les articles scientifiques, nous donnons une petite définition de la recherche scientifique, une définition de l'évaluation des articles scientifiques puis nous listons ses différentes techniques ainsi que les facteurs utilisés pour évaluer les articles. Nous terminons ce chapitre par une présentation des bases de données existantes.

1.2. Articles scientifiques

Les articles scientifiques [1] servent à partager une recherche initiale avec d'autres scientifiques ou à passer en revue les recherches d'autres chercheurs. En tant que tels, ils sont essentiels au développement de la science moderne, dans laquelle le travail d'un monde dépend d'un autre. Pour atteindre leur objectif, les journaux doivent viser les médias et non la persuasion. Ils doivent être lisibles, clairs, précis et concis.

1.3. La recherche scientifique

La recherche est systématique [2] dans la mesure où elle suit des étapes ordonnées de manière logique :

- Comprendre la nature du problème étudié et identifier les champs de connaissances en lien avec un tel problème ;
- Établir l'état de l'art, c'est-à-dire collecter/étudier la littérature pour comprendre comment les autres chercheurs ont approché le problème ;
- Collecter les données de manière organisée et contrôlée en vue d'arriver à des décisions valides.
- Analyser les données appropriées au problème étudié ;
- Tirer les conclusions qui s'imposent et faire les généralisations qu'il faut.

1.4. Évaluation des articles scientifiques [3]

Les revues scientifiques constituent le principal canal de publication dans de nombreux domaines, et la communauté scientifique internationale, par le biais de l'évaluation par des pairs, a développé un système de contrôle de qualité du contenu de ces revues. Bien que le système ne soit pas sans failles, le fait que les articles aient été relus et approuvés par deux experts (en général) dans le domaine constitue une évaluation positive externe des résultats de recherche. La même chose peut être dite à propos des rapports de conférence, rassemblés dans un livre ou un journal, même si on sait que l'examen par les pairs est dans ce cas-ci un peu moins strict. Dans tous les autres cas, les évaluateurs doivent être conscients de la nécessité d'essayer d'obtenir eux-mêmes une expertise de la qualité du manuscrit. Nous donnons quelques conseils pour ce faire ci-dessous.

La communauté scientifique internationale doit également prêter attention aux données quantitatives qui sont censées mesurer l'impact de l'article ou du journal dans lequel l'article est publié, données basées principalement sur le nombre de citations qu'un article ou la revue reçoit pour d'autres publications dans le monde entier.

Nous allons maintenant décrire certains éléments dont on doit tenir compte pour l'évaluation d'une publication scientifique. Le lecteur ne devrait avoir aucune difficulté à adapter les observations à d'autres formes de publication, chaque fois que le terme « papier », « manuscrit » ou « article » est utilisé.

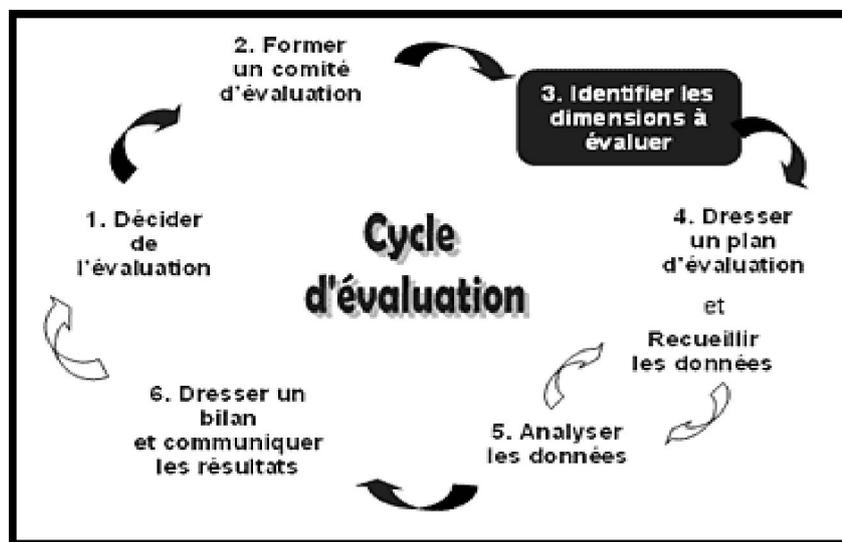


Figure 1 - Cycle d'évaluation

1.5. Techniques d'évaluation [5]

1.5.1. L'observation en situation

L'observation est plus qu'une techniques passive. Elle consiste à prélever des informations qualitative ou quantitative de manière systématiques. On observe des individus ou des groupes.

On peut observer de deux façons : en participant à la vie des sujets qu'on observe ou en demeurant en retrait de leur vie. On observe, pour ainsi dire, de l'intérieure ou de l'extérieur d'un groupe. On observe des comportements définis ou des profils de comportement qu'on cherche à circonscrire. On utilise le plus souvent des grilles d'observation encadrées par des définitions précises de comportements à observer. L'observateur ne recherche pas ce qui est insolite ou inusité : il quantifie plutôt des comportements bien définis, attendus, prévus.

On observe :

- ✓ Soit un individu en particulier
- ✓ Soit la communauté dans son ensemble

1.5.2. L'observation participante

L'observation participante est plus exigeante à bien des égards que l'observation désengagée. On tente de construire un portrait d'ensemble plus complexe et une compréhension plus profonde de la situation. Elle se caractérise surtout par l'établissement de relations entre les individus, leur rôle respectif au sein du groupe et le lien entre leur fonction et leur comportement général dans la communauté.

Pour retirer des informations pertinentes lors de sessions d'observation participante, vous utiliserez des techniques de recherche complémentaires qui visent la compréhension profonde du groupe étudié. Vous ferez appel, par exemple, à l'entrevue individuelle, à l'entrevue de groupe et à l'analyse de documents historiques.

Le danger de l'observation désengagée c'est le manque de profondeur, le manque de lien entre ce qu'on observe et sa signification.

L'observation désengagée est utile pour déterminer et comptabiliser des comportements. L'observation participante est plus efficace lorsqu'on veut obtenir sur une compréhension systémique.

1.5.3. L'entrevue de recherche

L'entrevue de recherche est une technique d'interrogation. On peut interroger un seul individu ou un groupe. Cette technique est souvent utilisée pour connaître un sujet encore inconnu ou pour approfondir un sujet méconnu. A travers ces rencontres surgiront des pistes de réflexion et des concepts qui pourront être soumis, par la suite, à une enquête menée auprès

de la population générale. Donc, à la suite de ces entrevues, on sera mieux en mesure de construire un questionnaire écrit plus complet et surtout, plus pertinent.

L'objectif de l'entrevue n'est pas d'amasser des collections d'idées disparates auprès d'un individu ou d'un groupe, mais plutôt de trouver des traits communs entre les idées exprimées. Il va de soi aussi que la science s'intéresse peu aux phénomènes singuliers, isolés : on recherche plutôt les phénomènes communs partagés par un ensemble d'individus. Il n'est donc pas étonnant de constater que les individus interrogés doivent entretenir des liens étroits avec le phénomène ou le sujet qu'on se propose d'étudier.

1.5.4. Le questionnaire ou le sondage

Le questionnaire est un outil qui permet de prélever des réponses de manière systématique. Comme la plupart des questions sont de type fermé, il s'ensuit que le questionnaire a été construit en fonction d'une conception ou d'une théorie précise, si bien que les réponses à choisir sont déjà toutes prêtes. On souhaite aussi que les possibilités de réponses soient exhaustives. Le questionnaire est une technique de recherche populaire, mais il n'est pas aussi simple à construire qu'il le semble. Pour la formulation des questions et pour en savoir davantage sur les qualités essentielles de la définition des variables, nous vous référons aux modules intitulés La mesure et Les variables.

L'objectif du questionnaire c'est de mesurer des aspects reliés à une problématique quelconque. Les questions ne sont pas posées au hasard. Elles s'inscrivent toutes dans des définitions opérationnelles des concepts prioritaires de la recherche.

En lisant un questionnaire, on discerne la plupart du temps les concepts privilégiés par la chercheuse ou le chercheur et la définition opérationnelle donnée à chacun d'eux. Il est bien évident que les chercheurs échangent entre eux des définitions opérationnelles et des concepts. Dans un domaine de recherche, des sections de questionnaire et même des questionnaires complets de retrouvent dans plusieurs recherches. En général, les concepts définis dans ces questionnaires sont considérés comme complets et d'excellente qualité.

C'est ainsi qu'en utilisant les mêmes concepts et les mêmes questions on est mieux en mesure de comparer les résultats d'une recherche à l'autre.

Un questionnaire expédié à un échantillon sous-entend qu'on nourrit un intérêt plus grand pour les données de groupe que celles de chaque individu.

- ✓ Les résultats individuels n'ont de l'importance que pour autant qu'ils contribuent à l'ensemble du groupe échantillonné.
- ✓ On vise à définir ce qui est constant et universel dans la population au moyen d'un échantillon représentatif.
- ✓ La qualité la plus importante que doit posséder un échantillon c'est la représentativité ou sa similitude avec la population générale.

1.5.5. L'expérimentation

L'expérimentation est une technique de recherche dont la caractéristique principale est le contrôle et la manipulation de variables. L'objectif ultime de ce contrôle strict est d'établir des liens de causalité ou de relation fonctionnelle. Cette technique est surtout utilisée en laboratoire. Inutile de mentionner que le contrôle strict de variables permet d'emblée de quantifier avec précision des variables et leurs interactions. Ce contrôle s'exerce plutôt aisément dans le cas de variables connues. Il y a deux types de contrôle de variables : le contrôle expérimental et le contrôle statistiques. Dans le cas du contrôle expérimental, une variable est contrôlée si elle est manipulée. On prend des moyens physiques, concrets et mécaniques pour équilibrer les sujets en fonction de critères jugés pertinents quant au sujet de recherche. Lorsqu'on pise au hasard des sujets pour un sondage d'opinion, c'est qu'on souhaite ainsi garder le même équilibre dans l'échantillon que dans la population.

Dans le cas du contrôle statistique ; une partie des données sont ajustées en fonction d'une ou plusieurs variables. L'ajustement est appliqué pour corriger un biais quelconque.

1.5.6. L'analyse de contenu

L'analyse de contenu est une technique de traitement de données. Cependant, ces données ne sont pas des nombres, mais plutôt des articles de journaux, des bandes sonores, des documentaires visuels, des archives historiques, des chansons et bien d'autres documents de ce genre.

Il s'agit de prélever systématiquement dans ces documents de divers types des informations de manière à mieux comprendre un phénomène. Le prélèvement est systématique parce qu'il se fait la plupart du temps à l'aide d'une grille quelconque qui même être modifiée en cours de route. Le prélèvement peut aussi bien être d'ordre qualitatif que quantitatif.

L'analyse de contenu permet d'interpréter des données qui n'ont pas été recueillies dans le but d'une recherche. Les productions ont été spontanées et faites dans une autre perspective. Un chercheur regroupe, par exemple, les écrits de journalistes portant sur un thème particulier. Ces journalistes n'ont pas écrit des textes en croyant qu'ils seraient analysés plus tard selon une grille particulière. Les techniques d'analyse traditionnelles ne permettent pas d'étudier ces écrits de journalistes, par exemple. L'analyse de contenu est très flexible quant à la matière brute dont elle fera usage. Elle part, bien souvent, de productions spontanées et de réalités indépendantes.

1.5.7. L'analyse statistiques

La réalisation et la rédaction d'une analyse statistique pour un article de recherche ou dans le simple but de rendre compte du travail effectué mettent en jeu deux domaines de compétence :

- ✓ Le calcul statistique avec ses méthodes, ses modèles, ses termes techniques et ses formules, leurs conditions d'applications, l'utilisation des logiciels ad hoc, les conclusions mathématiques licites auxquelles elles aboutissent ;
- ✓ L'écriture du rapport d'analyse via la présentation du protocole et des données (voire des hypothèses sous-jacentes), la mise en forme de tout ou partie des résultats et la rédaction des conclusions, interprétations et commentaires pour les spécialistes du domaine, pour l'équipe de recherche ou pour le grand public.

Une analyse statistique ne se réduit donc pas à une suite de calculs, même justes et justifiés. La rédaction est un art difficile. Elle est souvent bâclée par les [pseudo]scientifiques qui confondent phrases, littérature, verbiage et production littéraire. Or, la qualité d'un article de recherche, d'un rapport d'expérimentation transparait au fil des paragraphes. Le choix des termes employés renforce chez le lecteur ou le correcteur la conviction que le travail fait a été bien fait, que les méthodes statistiques sont maîtrisées, que le passage des chiffres (comme $m = 12.3$ jours) aux lettres.

1.5.8. La bibliométrie [6]

La bibliométrie est une méthode d'analyse quantitative utilisant les publications scientifiques et différents indicateurs pour mesurer la performance de la recherche, particulièrement 2 composantes importantes de la performance de la recherche:

- ✓ La **production scientifique**, soit la quantité de résultats scientifiques produits ;
- ✓ L'**impact scientifique**, soit l'influence des résultats scientifiques sur la progression subséquente de la science.

La bibliométrie permet d'évaluer :

- ✓ Les auteurs
- ✓ Les articles scientifiques
- ✓ Les revues
- ✓ Les institutions
- ✓ Les pays

1.5.8.1. Les bases de la bibliométrie

La bibliométrie est un ensemble de techniques visant à s'appuyer sur l'analyse des publications scientifiques pour mesurer la production de connaissances nouvelles – se sont véritablement développées à partir des années 1950, lorsque des chercheurs ont pu établir des statistiques sur la science à partir de bases de données bibliographiques recensant les publications scientifiques (livres, articles et communications écrites) du monde entier.

1.6. Les facteurs d'évaluation de la qualité d'un article scientifique [7]

Cette partie présente les principaux indicateurs des publications scientifiques : facteur d'impact ou FI, SC imago Journal Rank ou SJR, Eigen facteur, Source Normalized Impact per Paper ou SNIP, Article Influence, facteur h, et facteur g.

Nous allons classifier les différents facteurs selon trois classes principales : les facteurs associés à l'auteur, les facteurs associés à la publication et les facteurs associées à la revue scientifique.

1.6.1. Les facteurs de revues

1) Impact Factor (facteur d'impact - FI)

Le facteur d'impact d'une revue J en l'an X est le nombre de citations reçues en l'an X par l'ensemble des articles publiés dans la revue J au cours des 2 années précédentes, divisé par le nombre d'articles publiés par la revue J pendant la même période.

2) Scimago Journal Rank (SJR)

Nombre de fois où en moyenne un article dans une revue donnée est cité par d'autres articles, chaque citation étant pondérée par la notoriété de la revue citante.

Le SJR d'une revue J en l'an X est le nombre de citations reçues en l'an X par l'ensemble des articles publiés dans la revue J au cours des 3 années précédentes, pondérées par la notoriété des revues citantes, divisé par le nombre d'articles publiés par la revue J pendant la même période. $SJR_i = c. \frac{PSJR_i}{ART_i}$

3) Eigen Factor

L'Eigen factor d'une revue J en l'an X est le pourcentage de citations reçues par l'ensemble des articles de la revue J au cours des 5 années précédentes sur le total de citations reçues pendant la même période par l'ensemble des articles de l'ensemble des revues enregistrées dans le Journal Citation Reports (JCR).

4) Article Influence AI

L'influence d'un article d'une revue J en l'an X est le rapport entre l'Eigen factor et le nombre d'articles publiés par la revue J au cours des 5 années précédentes divisé par le nombre d'articles publiés pendant la même période dans l'ensemble des revues enregistrées dans le Journal Citation Reports (JCR).

5) Source Normalized Impact per Paper (SNIP)

Impact de citation d'une revue dans un champ disciplinaire donné. Un champ disciplinaire d'une revue J est défini par l'ensemble des articles qui citent la revue J et donc l'ensemble des revues auxquels appartiennent ces articles.

Le SNIP d'une revue J est le rapport entre le nombre de citations reçues par les articles de J publiés pendant les 3 années précédentes divisé par le nombre d'articles publiés par J pendant la même période, et le potentiel de citation dans le champ disciplinaire de la revue J. Le potentiel de citation indique la fréquence à laquelle les articles dans le champ disciplinaire d'une revue citent d'autres articles publiés dans les 3 années précédentes. Normalisé, une revue médiane dans son champ disciplinaire a un potentiel de citation de 1. La mesure est calculée comme suit :

Équation 1 : $SNIP = RIP / (R/M)$

où **RIP** = impact brut par papier, **R** = potentiel de citation et **M** = potentiel médian de citation de base de données.

1.6.2. Les facteurs pour les auteurs [8]

1) H index

Le h-index (ou facteur h), créé par le physicien Jorge Hirsch en 2005, est un indicateur d'impact des publications d'un chercheur. Il prend en compte le nombre de publications d'un chercheur et le nombre de leurs citations. Le h-index d'un auteur est égal au nombre h le plus élevé de ses publications qui ont reçu au moins h citations chacune.

Exemple : un h-index de 6 signifie que 6 publications de l'auteur ont chacune été citées au moins 6 fois

Le h-index est calculé en classant et en numérotant les publications de l'auteur de la plus citée (n° 1) à la moins citée. Le h-index correspond au dernier numéro de la publication qui vérifie : nombre de citations \geq numéro de la publication. Le h-index d'un chercheur est de plus en plus demandé dans les dossiers de soumission aux appels à projet ou dans le cadre d'évaluations de la recherche.

Le h-index peut aussi être calculé pour une entité de recherche (institution, laboratoire...), ou pour une revue. L'auteur considéré est le nom de l'entité (laboratoire, institution...) ou le titre de la revue.

2) G index (Facteur G)

Le facteur g est une variante du facteur h. Le facteur g est le plus grand nombre d'articles pour lesquels il est vrai de dire que l'ensemble des g articles ont reçu au moins g^2 citations.

Le facteur g met en avant les articles qui ont beaucoup plus de citations que le facteur h ne le laisse penser. Le facteur g est toujours supérieur au facteur h. $g^2 \leq \sum_{i \leq g} c_i$

3) HG index

Le hg-index est une combinaison des deux h-index et g-index. Il vise à fournir un plus équilibré vue de la production scientifique. Le hg-index d'un chercheur est défini comme la moyenne géométrique de ses indices h et g.

Équation 2: $hg - index = \sqrt{h * g}$

4) q^2 index

Cet indice fournit une vision plus globale de la production scientifique. Il est basé sur la moyenne géométrique de l'indice h, décrivant le nombre d'articles (dimension quantitative), et de l'indice m, décrivant l'impact des articles (dimension qualitative).

$$\text{Équation 3 : } q^2 - \text{index} = \sqrt{h * m}$$

5) H_r index

L'index h rationnel est une extension de l'index h d'origine. Il reflète le nombre de citations nécessaires pour augmenter l'indice h d'une unité. Mathématiquement :

$$\text{Équation 4: } h_r - \text{index} = (h + 1) - \frac{\text{Cit}(h+1)}{2h+1}$$

où Cit (h + 1) est le nombre de citations reçues par le (h + 1) - article.

6) H_i index

L'index h individuel est complémentaire à l'index h et estime le nombre d'articles qu'un chercheur aurait écrit tout au long de sa carrière avec au moins des citations s'il avait travaillé seul. La raison est de mesurer la productivité moyenne individuelle effective.

$$\text{Équation 5: } h_i - \text{index} = \frac{h}{N_a}$$

où N_a est le nombre moyen d'auteurs dans les articles sur h-core.

7) ORPI

Il indique l'originalité du chercheur par le biais de publications d'articles originaux par rapport au nombre total de publications au cours de la période considérée, quel que soit le classement de la revue. Il indique comment l'auteur se débrouille en termes de citations. Il annule le biais de l'autocitation qui se glisse autrement dans d'autres indices. Il élimine l'interférence de la paternité des cadeaux puisque la paternité des cadeaux n'est généralement pas reçue en tant que première paternité. Il indique la continuité des résultats de recherche originaux. Cela donne plus de poids au premier auteur et incite les chercheurs à inculquer l'habitude d'apporter une contribution originale à la recherche. Les articles de mérite originaux publiés dans des revues à faible FI obtiennent également un crédit. Ainsi, le score ORPI est un indicateur d'originalité, de productivité et de visibilité, sans biais de citation.

ORPI est calculée comme suit :

$$\text{Équation 6 : } ORPI = \frac{N}{I} (C - Sc) / T$$

Où ORPI est un acronyme pour Original Research Publication Index, en tant que premier auteur,

N = Nombre total d'articles originaux publiés dans des revues indexées Pubmed (puisque'il s'agit de la base de données la plus utilisée pour les citations) par l'auteur à partir de la première publication indexée jusqu'à la date,

C = Nombre total de citations reçues par «N» articles originaux publiés dans une revue indexée par l'auteur à partir de la première publication indexée jusqu'à la date,

SC = Nombre total d'auto-citations sur le «N» articles originaux publiés dans une revue indexée par l'auteur à partir de la première publication indexée jusqu'à la date,

I = Nombre total d'éléments citables, c'est-à-dire articles originaux, revues, rapports de cas, actes de procédure publiés par l'auteur dans une revue indexée par l'auteur à partir de la première publication indexée jusqu'à la date,

T = Durée en années à partir de la première publication indexée jusqu'à la date (cela donnera la profondeur temporelle de la piste de publication).

2) A index

L'indice a est défini comme le nombre moyen de citations reçues par les articles inclus dans le noyau h, c'est-à-dire les premiers articles. Cet indice mesure l'intensité de citation des articles h-core; Cependant, il peut être très sensible à quelques articles recevant un nombre de citations élevé.

4) c-index

Cet indice mesure la créativité, définie comme la génération de nouvelles connaissances scientifiques. Son but est de mettre en évidence des articles qui reçoivent de nombreuses citations et qui ont peu de références bibliographiques. Cet index est calculé à partir de la liste des citations et références d'articles de l'auteur.

$$\text{Équation 7 : } c - \text{index} = \sum_{i=1}^{N_p} \frac{c(n_i, m_i)}{a_i}$$

Où : $c(n_i, m_i) = m_i - n_i + \frac{n_i}{Ae^{az} + Be^{bz}}$; N_p est le nombre total d'articles publiés ; n_i est le nombre de références de l'article i ; m_i est le nombre de citations de l'article ; a_i est le nombre d'auteurs de l'article i ; $z = (m_i - 1)/(n_i + 5)$ et A, B, a, b sont des paramètres arbitraires .

8) M quotient(m-index)

Pour faciliter les comparaisons entre les académiciens ayant différentes durées de carrière académique, Hirsch a proposé une mesure «m » qui est obtenue en divisant l'indice h par le nombre d'années d'activité de l'académicien (mesuré depuis le premier article publié). Cela discrimine les académiciens qui travaillent à temps partiel ou qui ont eu une interruption de carrière. Il est donné par la formule :

$m = h \text{ index} / \text{nombre d'années depuis que l'académicien est actif depuis le premier article publié.}$

L'indice m permet des comparaisons entre les académiciens qui ont eu différentes durées de carrière universitaire et ceux qui ont eu une ou plusieurs interruptions de carrière au cours de leur carrière universitaire.

1.6.3. Les facteurs de la publication scientifique

1) Langue de la publication

La présomption selon laquelle la langue d'un article affecte son évaluation est justifiée par la littérature précédente. L'anglais étant la langue la plus utilisée dans le monde scientifique,

les éditeurs de revues préfèrent attirer un plus grand nombre de lecteurs, ce qui leur donnera plus de visibilité, plus de citations et plus de FI, comparé à l'allemand, le latin et le grec, qui étaient autrefois des langues populaires.

2) Nombre de citation

C'est un indicateur qui permet de mesurer l'intérêt d'un article en fonction de l'impact des citations. Cependant, un chercheur peut citer ses propres travaux, augmentant ainsi le nombre de citations qui lui seront créditées. C'est ce qu'on appelle le phénomène de « auto-citations » qui justifie les critiques sur la fiabilité de cette mesure.

1.7. Les bases de données bibliographiques [9]

La bibliométrie a pour objet de mesurer la visibilité ou l'impact des écrits scientifiques après leur parution en tant que publications, grâce à l'étude des documents tels qu'ils sont indexés dans les bases de données bibliographiques. Alfred Lotka, Samuel Bradford ou George Kingsley Zipf énoncent leurs lois entre 1926 et 1936. Eugène Garfield, avec son Index de citations en 1963 (Citation indexes for science : a new dimension in documentation through association of ideas) affine cette science en développant la scientométrie. Il utilise les citations présentes dans les articles scientifiques, c'est-à-dire les renvois faits à d'autres articles, pour lier les articles entre eux. Eugène Garfield fonda ISI, qui a connu un très gros succès et a été racheté par Thomson Reuters.

Aujourd'hui, Thomson Reuters (ISI : Web of Science) possède l'une des trois principales bases bibliométriques au monde. Les deux autres appartiennent à des acteurs privés : **Elsevier (Scopus)** et **Google (Google scholar)**.

Nous vous les présentons ci-après avec une attention particulière pour **Google Scholar**, d'accès libre. Nous verrons ensuite **WorldCat**, plus grand catalogue de bibliothèques au monde, ainsi que la base de données **BASE**, toutes deux également libres mais qui émanent d'organisations à buts non lucratifs.

1.7.1. Thomson Reuters (ISI - Institute for Scientific Information)

L'Institute for Scientific Information (ISI) créé en 1960 par Eugene Garfield a été acquis par Thomson Scientific & Healthcare en 1992 et a porté le nom de Thomson ISI. À la suite de la fusion avec Reuters en 2008, la compagnie fait maintenant partie de la Thomson Reuters Corporation.

L'ISI a développé son activité autour de sept bases de données de publications scientifiques et de leurs citations via Web of science, une interface web qui permet d'analyser les citations d'une personne, d'un groupe, d'un article ou d'un journal un service d'information universitaire en ligne.

Web of Science, un service d'information universitaire en ligne qui donne accès à sept bases de données bibliographiques :

- Conference Proceedings Citation Index,
- Science Citation Index Expanded,
- Social Sciences Citation Index,
- Arts & Humanities Citation Index,
- Index Chemicus,
- Current Chemical Reactions,
- Conference Proceedings Citation Index: Science and Social Science and Humanities.

1.7.2. Elsevier (Scopus)

Scopus est la plus grande base de données de résumés et de citations de la littérature évaluée par des pairs : revues scientifiques, livres et actes de conférences. Présentant un aperçu complet des résultats de la recherche mondiale dans les domaines de la science, de la technologie, de la médecine, des sciences sociales et des arts et sciences humaines, Scopus propose des outils intelligents pour suivre, analyser et visualiser la recherche.

1.7.3. Google Scholar [11]

Google Scholar, lancé fin 2004, permet d'effectuer des recherches sur des articles scientifiques approuvés ou non par des comités de relecture (peer-reviewed), des thèses de type universitaires, citations ou encore des livres scientifiques.

L'index de Google Scholar contient la plupart des journaux en ligne soumis à des comités de relectures, journaux provenant des grands éditeurs de littérature scientifique. Cependant la couverture réelle n'est pas connue précisément et plusieurs chercheurs ont repéré des trous dans sa couverture des archives des éditeurs.

La qualité des résultats des recherches n'atteint pas celle des bases de données classiques. Bien que l'attention ait été attirée sur le fait que Google Scholar n'est pas un outil adapté pour des applications bibliométriques ni pour l'évaluation de l'impact des chercheurs et des revues, sa simplicité d'utilisation et sa gratuité en font un outil de recherche populaire.

Google Scholar n'est pas le seul à permettre aux universitaires, aux chercheurs et aux ingénieurs de dénicher la perle rare sur le Net.

Il existe aussi plusieurs moteurs de recherche spécialisés dans l'information scientifique, de bonne facture et très performants. Les moteurs de recherche scientifique permettent non seulement d'accéder aux métadonnées d'une revue, mais souvent à tous les articles de ses numéros, en liaison avec diverse bases de données de littérature scientifique.

1.7.4. WorldCat

Créé en 1971, WorlCat est la contraction de l'anglais World Catalog (catalogue mondial). WorldCat est en accès libre via le Web sur WorldCat.org depuis 2006.

WorldCat rassemble les catalogues de plus de 10 000 bibliothèques publiques et privées à travers le monde. C'est la base de données bibliographiques en ligne de l'OCLC (Online Computer Library Center) qui est une organisation à but non lucratif mondiale, créée en 1967, au service des bibliothèques, avec pour but d'offrir un meilleur accès public aux informations et d'en réduire le coût. Il est réputé être le plus grand catalogue OPAC (Online public Access catalog) du monde.

Dix ans plus tard, son moteur de recherche donne accès à plus de deux milliards d'ouvrages et permet d'obtenir des références bibliographiques dans les cinq styles les plus utilisés et d'exporter ces références.

Concernant la recherche d'article, il est possible de limiter à certaines revues, périodiques qu'il vous faudra spécifier.

1.7.5. BASE

BASE (Bielefeld Academic Search Engine) est un moteur de recherche développé dans le cadre du projet Open Archives Initiative (OAI) par l'Université de Bielefeld (Allemagne). BASE est un des plus importants moteurs de recherche au monde consacré aux publications universitaires et scientifiques.

Il se distingue d'autres moteurs de recherche par sa spécialisation universitaire, la capacité de trouver des ressources appartenant au Web profond (ou Web invisible, non indexé par les moteurs commerciaux), l'utilisation de la recherche plein texte et la présentation des résultats de recherche accompagnés des données bibliographiques.

Il donne accès à plus 87 832 050 documents provenant de 4028 sources de contenus (à la date de rédaction de l'article). Plus de trois quarts d'entre eux sont proposés dans leur intégralité. Beaucoup d'universités françaises y sont représentées. BASE effectue l'indexation automatique des bibliothèques numériques utilisant le protocole Open Archives Initiative Protocol for Meta data Harvesting (OAI-PMH).

1.7.6. Semantic Scholar

Semantic Scholar est un projet développé par l'Institut Allen pour l'intelligence artificielle. Paru publiquement en novembre 2015, il est conçu pour être un moteur de recherche d'articles de revues scientifiques basé sur l'IA. Le projet utilise une combinaison d'apprentissage automatique, de traitement du langage naturel et de vision industrielle pour ajouter une couche d'analyse sémantique aux méthodes traditionnelles d'analyse des citations et pour extraire des figures, des entités et des lieux pertinents des documents.

En janvier 2018, à la suite d'un projet de 2017 comprenant l'ajout d'articles sur le biomédical et de résumés de sujets, le corpus des érudits sémantiques comprend désormais plus de 40 millions d'articles sur l'informatique et la biomédecine.

1.8. Conclusion

Une nouvelle connaissance, une découverte scientifique, sont avérées à partir du moment où elles sont validées par la communauté scientifique. Cette validation se fait, généralement, via une publication scientifique ou une prise de brevet. La qualité scientifique d'un chercheur, d'un laboratoire, d'un organisme de recherche ou d'une partie de la communauté scientifique est donc appréciée via son évaluation scientifique.

Chapitre 2 :

Méthodes et stratégies d'évaluation

2.1. Introduction

Ce chapitre offre des directives pour l'utilisation de méthodes et techniques d'évaluation spécifiques des articles scientifiques.

Le choix d'une méthode d'évaluation dépend des objectifs de l'évaluation et il faut souligner que ces objectifs ne sont pas toujours les mêmes. On doit donc bien définir ce que l'on cherche à évaluer et sélectionner les méthodes ou techniques appropriées en fonction des données que l'on doit recueillir.

À l'intérieur de ce document, la description de chaque méthode comprend des remarques concernant les objectifs qu'elle permet d'évaluer et la situation dans laquelle on y ferait appel. Ces indications devraient faciliter le choix de méthodes et techniques appropriées.

2.2. Les techniques d'évaluation

Dans cette partie, nous allons présenter les diverses méthodes et techniques informatique pouvant servir à formaliser un problème et à l'évaluer et ceci afin de choisir l'approche et l'outil les plus adéquats pour résoudre notre problématique.

2.2.1. L'intelligence Artificielle

L'intelligence artificielle est un ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine. Autrement dit, c'est la concrétisation de la formule : "l'esprit dans la machine". Mélangeant mathématiques, informatique, robotique et sciences cognitives, c'est un champ qui recouvre de multiples applications, de la voiture autonome capable d'éviter un piéton au programme susceptible de reconnaître instantanément un visage dans une foule.

L'intelligence artificielle (IA) est un domaine de l'informatique qui met l'accent sur la création de machines intelligentes qui fonctionnent et réagissent comme les humains. Certaines des activités pour lesquelles les ordinateurs avec intelligence artificielle sont conçus incluent : Reconnaissance de la parole, Apprendre, Planification, Résolution de problèmes.

2.2.2. L'apprentissage Machine [12]

L'apprentissage machine est également au cœur de l'IA. L'apprentissage sans aucune forme de supervision exige la capacité d'identifier des modèles dans les flux d'intrants, alors que l'apprentissage avec une supervision adéquate implique une classification et des régressions numériques. La classification détermine la catégorie à laquelle appartient un objet et la régression traite de l'obtention d'un ensemble d'exemples d'entrées ou de sorties numériques, découvrant ainsi des fonctions permettant la génération de sorties appropriées à partir des entrées respectives. L'analyse mathématique des algorithmes d'apprentissage machine et de leurs performances est une branche bien définie de l'informatique théorique, souvent appelée théorie de l'apprentissage computationnel.

La perception machine traite de la capacité d'utiliser des entrées sensorielles pour déduire les différents aspects du monde, tandis que la vision par ordinateur est le pouvoir d'analyser les entrées visuelles avec quelques sous-problèmes tels que la reconnaissance faciale, la reconnaissance d'objets et de gestes.

Il est virtuellement impossible d'écrire un programme qui fonctionnera de manière robuste dans toutes les situations. C'est là qu'intervient l'apprentissage machine (que l'on appelle aussi apprentissage automatique). C'est l'apprentissage qui anime les systèmes de toutes les grandes entreprises d'Internet. Elles l'utilisent depuis longtemps pour filtrer les contenus indésirables, ordonner des réponses à une recherche, faire des recommandations, ou sélectionner les informations intéressantes pour chaque utilisateur.

Un système entraînable peut être vu comme une boîte noire avec une entrée, par exemple une image, un son, ou un texte, et une sortie qui peut représenter la catégorie de l'objet dans l'image, le mot prononcé, ou le sujet dont parle le texte. On parle alors de systèmes de classification ou de reconnaissance des formes.

Dans sa forme la plus utilisée, l'apprentissage machine est supervisé : on montre en entrée de la machine une photo d'un objet, par exemple une voiture, et on lui donne la sortie désirée pour une voiture. Puis on lui montre la photo d'un chien avec la sortie désirée pour un chien. Après chaque exemple, la machine ajuste ses paramètres internes de manière à rapprocher sa sortie de la sortie désirée. Après avoir montré à la machine des milliers ou des millions d'exemples étiquetés avec leur catégorie, la machine devient capable de classer correctement la plupart d'entre eux. Mais ce qui est plus intéressant, c'est qu'elle peut aussi classer correctement des images de voiture ou de chien qu'elle n'a jamais vues durant la phase l'apprentissage. C'est ce qu'on appelle la capacité de généralisation.

2.2.3. Les Algorithmes génétiques [14]

Les algorithmes génétiques (AGs) sont des algorithmes d'optimisation stochastique fondés sur les mécanismes de la sélection naturelle et de la génétique. Leur fonctionnement est extrêmement simple. On part avec une population de solutions potentielles (chromosomes) initiales arbitrairement choisies. On évalue leur performance (fitness) relative. Sur la base de

ces performances on crée une nouvelle population de solutions potentielles en utilisant des opérateurs évolutionnaires simples : la sélection, le croisement et la mutation. On recommence ce cycle jusqu'à ce que l'on trouve une solution satisfaisante.

Les AGs ont été initialement développés par John Holland (1975). Leurs champs d'application sont très vastes. Outre l'économie, ils sont utilisés pour l'optimisation de fonctions, en finance, en théorie du contrôle optimal, ou encore en théorie des jeux répétés et différentiels. La raison de ce grand nombre d'application est claire : simplicité et efficacité. Bien sûr d'autres techniques d'exploration stochastique existent, la plus connue étant le recuit simulé. La raison de ce grand nombre d'application est claire : simplicité et efficacité. Bien sûr d'autres techniques d'exploration stochastique existent, la plus connue étant le recuit simulé.

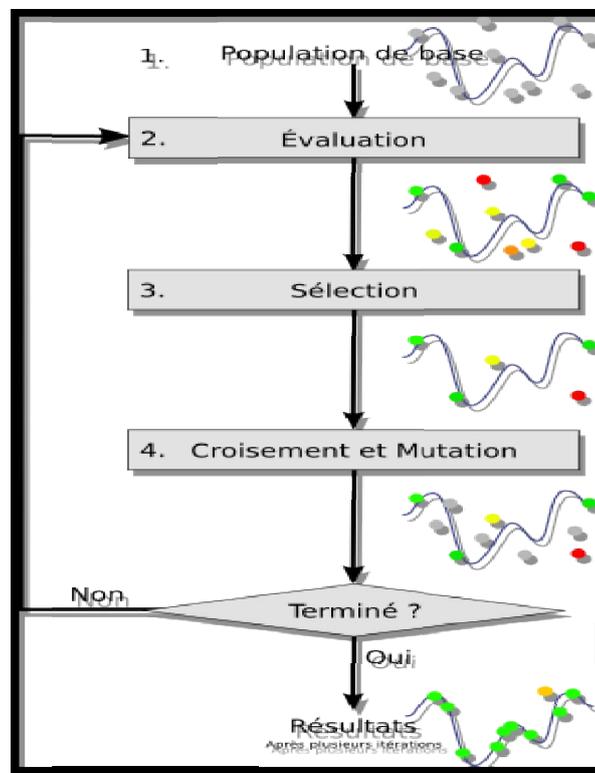


Figure 2 - Organigramme d'algorithme génétique

- La population de base (genèse est l'étape de la création d'une population aléatoire. C'est le point de départ de notre algorithme
- L'évaluation est l'analyse des individus pour analyser si une solution est disponible. Pour ceci, nous utilisons une fonction de coût, ou d'erreur, afin de définir le score d'adaptation des individus lors du processus de sélection.
- Nous effectuons une boucle tant que l'évaluation estime que la solution n'est pas optimale.

2.3. Travaux existants sur l'évaluation d'articles scientifiques

2.3.1. Facteur d'impact pondéré (Weighted Impact Factor WIFE) [15]

Dans le calcul du facteur d'impact, on utilise le nombre total de citations de revues, quel que soit le prestige des revues citées, cependant, des unités scientifiques (papier, chercheur, revue ou organisation scientifique) cités par des revues à facteur d'impact élevé ou des chercheurs avec un indice de Hirsch élevé sont plus importants que les objets cités par des revues sans facteur d'impact ou chercheur inconnu.

Dans ce travail, nous allons présenter un facteur d'impact pondéré qui a été proposé pour obtenir plus un classement plus précis des revues, qui tient en compte non seulement la quantité de citations, mais également de la qualité des revues de citations.

2.3.1.1. Présentation de la stratégie

Supposons que S^t soit l'ensemble de toutes les revues indexées au cours de l'année t , et n^t le nombre de revues indexées, n_j^t est le nombre de journaux dans les articles de journaux j cités en S^t , a_j^t est le nombre total d'articles publiés dans l'année t , c_{ij}^t est le nombre de citations au journal j du journal i de l'année t , FIF_j^t est le facteur d'impact sur 5 ans du journal j au cours de l'année t . Le FI pondéré proposé pour le journal j au cours de l'année t est :

$$\text{Équation 8 : } WIF_j^t = \frac{\sum_{i=1}^{n_j^t} (FIF_j^{t-1} + 1) c_{ij}^t}{a_j^{t-1} + a_j^{t-2}}$$

S. No.	Journal	Number of articles published in 2011 and 2012	Number of citations in 2013 to articles published in 2011 and 2012	Number of cited journals with IF	Impact factor of journal in previous year
1.	Information Systems Journal	54	116	10	-
2.	Automated Software Engineering	32	68	14	1.4
3.	Knowledge Engineering Review	50	120	22	0.59

Tableau 1. Quelques indicateurs du journal analysé reçu de la base de données Scopus

Pour illustrer le fonctionnement du facteur d'impact de ce travail, nous présentons un exemple pour le calcul des facteurs d'impact pondéré pour une trois revues dans le domaine de l'informatique, avec des facteurs d'impact différents de Journal Citation Reports (JCR) (Tableau 1). Pour ces revues, le nombre d'articles publiés en 2011 et 2012, le nombre total de citations en 2013, le nombre de sources d'une revue de citations de facteurs d'impact a été déterminé à l'aide de données sur les articles et leurs citations tirées de la base de données Scopus. Le nombre de citations de chaque revue avec des facteurs d'impact a été déterminé pour des revues données (Tableau 2).

J1 - Information Systems Journal													
1	2	3	4	5	6	7	8	9	10				
3.61 (1)	3.53 (1)	2.04 (1)	1.77 (3)	1.66 (2)	1.52 (1)	1.27 (1)	1.13 (2)	0.86 (1)	0.6 (1)				
J2 – Automated Software Engineering													
1	2	3	4	5	6	7	8	9	10	11	12	13	14
4.63(1)	4.10(1)	2.59(1)	2.51(1)	1.85(1)	1.73(1)	1.52(1)	1.52(1)	1.4(3)	1.23(1)	1.14(1)	0.71(1)	0.49(1)	0.3(1)
J3 - Knowledge Engineering Review													
1	2	3	4	5	6	7	8	9	10	11	12	13	14
5.06(1)	4.87(1)	3.73(2)	3.64(1)	3.48(1)	2.51(1)	2.2(2)	2.17(1)	1.8(2)	1.57(1)	1.56(1)	1.42(1)	1.38(1)	1.38(1)

Tableau 2 : Facteurs d'impact de la citation de journaux et du nombre de leurs citations.

S. No.	Journal	JCR IF	Proposed WIF
1	Information Systems Journal	1.333 [1]	2.599 [6]
2	Automated Software Engineering	1.24 [2]	3.016 [5]
3	Knowledge Engineering Review	0.957 [3]	3.484 [3]

Tableau 3 :Facteurs d'impact des RJC et facteurs d'impact pondérés proposés dans les journaux analysés.

En conséquence, nous avons deux listes de classement différentes pour les revues analysées (tableau 3). Examinons deux revues avec des indicateurs similaires (nombre de citations, nombre de ressources de citations, nombre d'articles) - Information Systems Journal et Knowledge Engineering Review. Dans la liste de classement par proposition, le classement pondéré des facteurs d'impact de l'Information Systems Journal est passé du 1er au 6e rang. Malgré un nombre similaire de citations et d'articles publiés, le facteur d'impact du Knowledge Engineering Review Journal est plus élevé que celui de l'Information Systems Journal. Il s'explique par les facteurs d'impact élevés des revues de citations et un grand nombre de sources de citations prestigieuses de la revue Knowledge Engineering Review.

2.3.1.2. Les points forts de ce travail

Le facteur d'impact pondéré obtient un classement plus précis des revues, qui prennent en compte non seulement le nombre de citations, mais aussi la qualité des revues citées. Comme cas particulier de l'équation du facteur d'impact pondéré proposé, nous obtenons l'équation actuelle du facteur d'impact calculée par l'Institut d'information scientifique pour les RJC, mais le fait de considérer le prestige et l'influence des revues citées donne un indicateur plus précis du classement des revues. La méthode suggérée pour la FI peut également s'appliquer à l'évaluation d'un chercheur, d'une organisation ou d'un pays. Par conséquent, en ce qui concerne les index des objets cités pour l'évaluation de la recherche, les résultats sont plus efficaces.

2.3.2. Médias sociaux et altimétriques [18]

Une autre proposition d'évaluation consiste à mesurer la communication scientifique à travers divers canaux en ligne.

Les mesures altimétriques consistent à utiliser des méthodes alternatives pour mesurer la qualité d'un travail. Des méthodes telles que l'utilisation des réseaux sociaux.

Par exemple, les bibliothécaires sont encouragés à se tenir au courant des nouvelles statistiques pour mieux aider leurs clients à trouver des publications spécifiques. En outre, les éditeurs, ainsi que certaines sociétés de médias sociaux, s'intéressent également à l'altmetrics. En 2013, l'éditeur Elsevier a acheté la plate-forme de bookmarking social Mendeley.

En tant que l'un des principaux fournisseurs de métriques de citations, Elsevier semble être particulièrement intéressé par les développements en cours dans altmetrics.

2.3.2.1. Présentation

Les recherches actuelles sur l'altimétrie en sont encore à un stade précoce, mais elles se développent rapidement. L'histoire de la recherche altimétrique est également résumée par Fenner. Des érudits d'origines différentes jouent le rôle de pionniers dans le développement de nouvelles métriques et de nouvelles approches de recherche. La plupart des recherches actuelles visent à définir la qualité et la portée des indicateurs altmetrics.

La comparaison la plus complète des altimétriques et des citations à ce jour a été réalisée par Thelwall et al. Ont trouvé des mesures du lectorat de Mendeley pour 62,6% de toutes les publications de leur échantillon de test, et Priem et al. Ont constaté que près de 80% de leurs publications étaient incluses dans Mendeley. Trouvé que seulement environ 5% de leur échantillon de documents cités sur Wikipedia et Shuai et al. Les différences disciplinaires semblent sous-tendre la plupart des altimétriques actuels. Tous ces exemples montrent qu'il reste encore beaucoup à étudier avant de comprendre pleinement quelles métriques alternatives issues des plates-formes de médias sociaux ou d'autres données en ligne peuvent nous renseigner sur les activités de recherche et leur impact scientifique.

2.3.2.2. Les avantages et les inconvénients

Les recherches actuelles se concentrent sur cette tâche et créent, étape par étape, une carte du paysage altimétriques. Ceci n'est bien sûr pas nouveau et se produit également avec la bibliométrie traditionnelle et le décompte de citations, par exemple, à travers des autocitations ou des cartels. Mais les bibliométriciens ont trouvé des moyens d'ajuster leurs indicateurs afin de respecter un tel comportement. Par conséquent, des recherches indépendantes devraient surveiller l'utilisation des altimétriques avec autant de soin et continuer à étudier l'interaction des outils, des utilisateurs et des métriques.

Les altimétriques, sont des méthodes d'évaluation d'activités qui servent d'alternatives aux métriques basées sur les citations, et constituent un domaine de recherche relativement nouveau mais en croissance rapide.

2.3.3. Les Algorithmes génétiques et les Réseaux gaussiens bayésiens [8] [17]

2.3.3.1. Présentation des algorithmes génétiques (AGs)

Un algorithme génétique est défini par :

- ✓ Individu/chromosome/séquence : une solution potentielle du problème ;
- ✓ Population : un ensemble de chromosomes ou de points de l'espace de recherche ;
- ✓ Environnement : l'espace de recherche ;
- ✓ Fonction de fitness : la fonction - positive - que nous cherchons à maximiser.

2.3.3.2. Le principe des algorithmes génétiques

Les algorithmes génétiques utilisent la théorie de Darwin sur l'évolution des espèces. Elle repose sur trois principes :

- Le principe de variation : Chaque individu au sein d'une population est unique. Ces différences, plus ou moins importantes, vont être décisives dans le processus de sélection.
- Le principe d'adaptation : Les individus les plus adaptés à leur environnement atteignent plus facilement l'âge adulte. Ceux ayant une meilleure capacité de survie pourront donc se reproduire davantage.
- Le principe d'hérédité : Les caractéristiques des individus doivent être héréditaires pour pouvoir être transmises à leur descendance. Ce mécanisme permettra de faire évoluer l'espèce pour partager les caractéristiques avantageuses à sa survie.

2.3.3.3. Opérateurs d'évolution

Il y a trois opérateurs d'évolution dans les algorithmes génétiques :

- **La sélection** : Choix des individus les mieux adaptés.
- **Le croisement** : Mélange par la reproduction des particularités des individus choisis.
- **La mutation** : Altération aléatoire des particularités d'un individu.

2.3.3.4. Apprentissage de la structure BN (Bayesian Networks)

Un **réseau gaussien bayésien** est un réseau bayésien dont toutes les variables sont continuées et tous les CPD sont des Gaussiennes linéaires. Un réseau bayésien (BN) est un graphe orienté qui représente distribution de probabilité conjointe parmi un grand nombre de variables et permet de réaliser une inférence probabiliste avec ces variables. Il a été appliqué à un large éventail de tâches tels que les systèmes de dialogue parlés naturels, la reconnaissance de la vision, systèmes experts, diagnostic médical et inférence de réseau de réglementation génétique pour en nommer quelques-uns.

Un BN contient deux composants importants : un graphe acyclique dirigé (DAG) représentant la structure de dépendance entre les variables du réseau et un tableau de

probabilité conditionnelle (CPT) pour chaque variable du réseau, en fonction de son ensemble parent. Apprendre la structure de ces réseaux à partir de données est l'un des problèmes les plus difficiles.

Équation 9: $P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \Pi(x_i))$

Où : $\Pi(x_i)$ représente l'ensemble des parents de X_i .

2.3.3.5. Apprentissage GBNs en Utilisant AGs

Les GA et les GBN sont utilisés dans ce travail pour découvrir les sous-ensemble d'indices bibliométriques avec le plus fort prédictif pouvoir de tous. Une analyse d'emballage pour évaluer toutes les différentes structures est utilisée pour atteindre cet objectif. Par conséquent, après mise en place d'un découpage spécifique entre prédictif et réponse variables, nous utilisons un GA pour rechercher la structure optimale de GBN, qui minimise la distance entre réel et prédit valeurs de variable de réponse. Le processus est répété pour tous configurations possibles des prédicteurs et des nœuds de réponse.

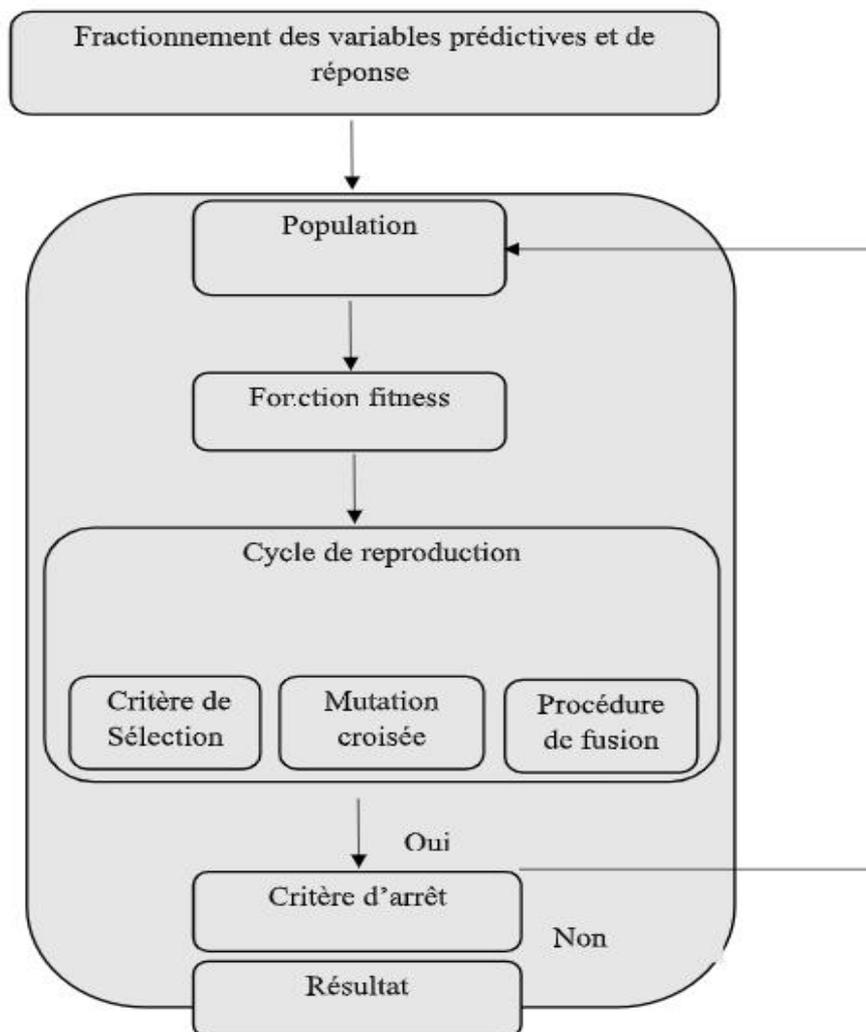


Figure 3 - Les étapes d'un algorithme génétique. [17] [8]

2.2.3.6. Explications

1) Population

L'espace de recherche des solutions candidates est représenté par une collection de N individus, appelée population. Dans notre problème, les individus représentent les structures GBN. Chaque structure est décrite par une matrice d'adjacence Adj. (G), qui est la représentation du graphe $G = (V(G), A(G))$. La matrice d'adjacence est une matrice $n \times n$ avec les entrées a_{ij} , $i, j = 1, \dots, n$, tel que $a_{ij} = 1$ si, et seulement si, un arc existe entre les nœuds i et j , et $a_{ij} = 0$ sinon. En utilisant cette codification, un individu peut être transformé en une chaîne binaire ($a_{11}, \dots, a_{1n}, a_{21}, \dots, a_{2n}, \dots, a_{n1}, \dots, a_{nn}$), qui transforme son contiguïté matrice sous forme vectorisée.

La population initiale est générée aléatoirement. Arcs dans les matrices d'adjacence sont tirées au hasard à partir d'une Bernoulli distribution avec $p = .5$ (probabilité de succès). Le cas échéant, la structure du réseau est modifiée pour éviter la présence de cycles.

2) La fonction fitness

Une fonction de mise en forme idéale doit être étroitement liée à l'objectif et doit être calculé rapidement. L'AG doit être itéré plusieurs fois pour produire des résultats fiables. Dans notre étude, étant donné un individu avec p prédictif les variables et les variables de réponse r , nous calculons la Mahalanobis distances entre les valeurs réelles et prédites pour les variables de réponse r .

L'objectif est de minimiser la distance entre les valeurs de variables de réponse réelles et prédites. En forme de :

$$\text{Équation 10: } MD(y, y') = \sqrt{(y, y')^T \Sigma_{yy}^{-1} (y, y')}$$

3) Cycle de reproduction

Les critères de sélection des parents, les opérateurs de croisement et de mutation, ainsi que la procédure de fusion sont les composants du cycle de reproduction dans une AG.

4) Critère de sélection

Le processus de sélection détermine lequel des individus de la population actuelle sera mate pour créer de nouveaux individus.

Cette stratégie garantit l'amélioration de la valeur moyenne et minimale dans chaque AG itération. Ainsi, les meilleurs $N / 2$ individus de la population pour la reproduction sont identifiés et ensuite déplacés dans piscine d'accouplement où ils sont combinés par croisement et mutation opérations.

5) Crossover et mutation

Dans le croisement, $N / 2$ parents sont accouplés au hasard deux à deux pour créer $N / 2$ nouveaux enfants en combinant leurs informations génotypiques. L'objectif du crossover est de produire des individus en forme en échangeant des informations contenues dans des individus déjà bons. Nous avons choisi l'opérateur de croisement à un point, qui est

l'opérateur le plus commun et donne de bons résultats. Compte tenu de la binaire codification des individus, on choisit au hasard avec une probabilité fixe P_c un point de croisement auquel les informations sont échangées.

La première progéniture prend la première section du premier parent et la dernière partie à partir de la deuxième, alors que la deuxième progéniture est formée inversement. L'opérateur de mutation introduit une variabilité supplémentaire dans la population pour améliorer le degré de diversité. Ces changements peuvent, à leur tour, entraîner de nouvelles personnes avec des scores de fitness plus élevés. Enfin, chaque fois qu'une progéniture viole la contrainte de graphe acyclique dirigée, l'opérateur supprime au hasard des arcs à modifier cycles.

6) Procédure de fusion

La dernière étape de la reproduction cycle est la génération de la nouvelle population. Encore une fois, nous avons choisi une stratégie élitiste pour générer la nouvelle population en combinant les meilleurs individus des générations précédentes et nouvelles. Le principal avantage de cette stratégie est qu'elle préserve toujours le meilleur sous-ensemble d'individus à chaque génération.

7) Critères d'arrêt

La recherche est arrêtée lorsqu'un ensemble de conditions, les critères d'arrêt, sont remplies. Différents critères pour arrêter une AG ont été développés dans la littérature : après un nombre spécifique de générations ou un nombre maximum d'évaluations, s'il n'y a pas d'amélioration de la fonction objectif, ou lorsque la fonction objectif délivre une valeur spécifique, entre autres. Ici, un nombre maximum de générations ou aucune amélioration sur un nombre donné de générations constitue notre critère d'arrêt. L'individu avec le score le plus élevé dans la population finale est considéré comme la solution au problème d'optimisation.

2.3.3.7. Les points fort et faible

De nombreux index bibliométriques ont été développés afin de prendre en compte des aspects précédemment découverts. Dans ce contexte, certains chercheurs ont récemment porté leur attention sur le pouvoir prédictif des indices bibliométriques dans de nombreuses situations. Le résultat est que la communauté scientifique est maintenant confrontée au défi de sélectionner parmi ce groupe d'indices bibliométriques ceux qui ont un pouvoir prédictif plus élevé.

2.4. Conclusion

Dans ce chapitre, nous avons étudié les techniques et les stratégies existantes qu'on peut utiliser pour évaluer les articles scientifiques, nous avons cité le cas d'utilisation pour chaque technique. Aussi, nous avons abordé quelques travaux connexes à notre projet.

Chapitre 3 :

Conception et développement

3.1. Notations

hw :	h-index weighted (Pondéré).	OPRI:	Original Research Publication Index.
r_w :	Weighted rank	n_j^t :	le nombre de revues dans l'ensemble de toutes les revues indexées dans l'année t pour citer les articles de revues
y_j :	le nombre de citation reçues par l'articles j .	c_{ij}^t :	le nombre de citations au journal j du journal i dans l'année t
r_0 :	e plus grand index de ligne i $Tq : r_w(j) \leq y_j$.	a_j^t :	le nombre total d'article publiés au cours de l'année t
h :	h-index.	A^t :	le nombre total des articles publiés par des revues de tout domaine scientifique donné de l'année en cours.
g :	g-index.	c^t :	le nombre total de citations pour l'ensemble du domaine scientifiques donné dans l'année en cours.
c	c-index	j^t :	le nombre de revues de la discipline particulière publiées.
	$C(n_i, m_i) = m_i - n_i + \frac{n_i}{Ae^{az} + Be^{bz}}$	\hat{C}^t :	le nombre de citations à la revue de discipline particulière.
N_p :	le nombre total d'articles publiés.	$SNIP$:	Source Normalized Impact per Paper.
n_i :	le nombre de références de l'article i .	RIP :	Raw Impact per Paper.
m_i :	le nombre de citation de l'article i .	R :	potentiel de citation.
a_i :	le nombre d'auteurs de l'article i .	M :	potentiel médian de citation de base de données.
z :	$z = (m_i - 1)/(n_i + 5)$	SJR_i :	SCImag journal Rank of the Journal
A, B, a, b :	sont des paramètres arbitraires	$PSJR_i$:	The Prestige SJR.

$m :$	h index / le nombre d'années depuis que l'académiciens est actif depuis le premier article publié.	$ART_i :$	Number of primary items (articles, reviews, and conference papers) of journal i.
$q^2 :$	q^2 -index.	$h_r :$	une extension de l'index h d'origine.
$h_i :$	hi-index	Hg :	Hg-index

3.2. Introduction

Ce chapitre porte sur la conception et le développement de notre solution. Nous présentons le contexte et motivation du projet, en présentant d'abord la problématique à résoudre et une analyse de l'existant qui résume les travaux étudiés dans le chapitre 2.

Ensuite, nous détaillons les démarches et étapes de résolution du problème posé. Enfin, nous décrivons les différents schémas et diagrammes de modélisation de la solution proposée.

3.3. Contexte et motivation du projet

3.3.1. Problématique

La recherche scientifique est essentielle au développement et progrès des nations. L'idée principale de notre travail est de faciliter l'étape de l'étude et de l'état de l'art. A cet effet, nous avons conçu un outil permanentant de donner une idée générale sur les références bibliographiques d'une publication scientifique et produit un score final pour chaque référence.

Ces différents scores donneront à l'utilisateur un aperçu sur la publication qu'il est entrain de consulter. Il pourra ainsi soit étudier la publication plus en détails si les scores de ses références sont concluant, soit décider de passer à une autre publication.

En dernier, l'outil proposé ne prétend pas donner un jugement final sur la publication, mais représente plutôt une aide pour le chercheur afin d'avoir une estimation générale sur les références de l'article et accélérer ainsi ses travaux de recherche.

3.3.2. Analyse de l'existant

Nous ne saurions débiter ce travail sans avoir une idée claire et précise sur l'existant quel qu'il soit. La première tâche à effectuer a été de commencer le travail en menant diverses recherches sur les travaux existants (Chapitre 2). Cette étape, nous a permis d'avoir une connaissance globale sur le sujet abordé et des solutions existantes.

Après avoir étudié le processus d'évaluation des articles scientifiques, les différents concepts et notions de la bibliométrie, ainsi que les techniques et méthodes de calcul des indicateurs scientifiques, nous avons pu déterminer la portée du projet étudié.

Les différents travaux présentés dans le chapitre précédent proposent principalement des solutions qui permettent d'évaluer les références en se basant sur un seul indicateur.

Ensuite, ces travaux ne prennent en compte qu'un seul aspect des références (en se focalisant soit sur l'article et ses indicateurs, soit sur l'auteur et ses indicateurs, soit sur le journal et ses indicateurs).

De plus, la plupart des travaux que nous avons cité restent des travaux théoriques et ne proposent pas un outil implémenté et utilisé en pratique.

Enfin, l'outil qui a été déjà implémenté [9], ne réalise qu'une simple moyenne des différents facteurs sans prendre en compte l'importance et particularité de chaque facteur.

3.3.3. Solution proposée

Face aux difficultés rencontrées dans l'évaluation des articles scientifiques sur des facteurs bibliométriques, nous avons pensé à améliorer un outil qui permet de calculer la qualité de référencement des publications académique et de les évaluer.

Suite, à l'analyse des travaux existants, notre proposition pour l'amélioration de l'outil d'évaluation des références consiste en :

1. Regrouper les facteurs du chapitre 2 en trois populations : auteur, revue, article, ce qui permet d'avoir plus de détails sur la référence et d'avoir des facteurs homogènes entre eux.
2. Associer à chaque facteurs un poids représentant son importance et utilité :
 - Les poids initiaux sont générés en calculant la corrélation entre l'article et un ensemble d'article de même catégories.
 - Les poids finaux associés à un facteur sont estimés en utilisant un algorithme génétique pour prédire le meilleur score d'évaluation.

Dans ce qui suit nous allons détailler les approches et principes de la solution proposée.

3.4. Approche de conception de la solution proposée

Dans la section précédente, nous avons présenté les motivations de notre travail. Nous avons aussi justifié le choix d'associer un poids à chaque facteur pour le calcul du score de chaque référence et ceci afin d'attribuer à chaque facteur son importance.

Ce poids est d'abord calculé en faisant la corrélation entre un article et d'autres articles de même catégorie. Mais, cette valeur ne correspond pas forcément à l'importance du facteur. Ainsi, nous avons pensé à utiliser un algorithme génétique afin d'estimer la meilleure valeur du poids.

Nous, avons utilisé l'algorithme génétique, car les méthodes de résolutions exactes, ne permettent pas de trouver la solution optimale dans une durée de temps raisonnable, contrairement aux métas heuristiques qui permettent de résoudre des problèmes d'optimisations complexes face auxquels les méthodes classiques manquent d'efficacité.

3.4.1. Algorithme génétique

Les algorithmes génétiques appartiennent à la famille des algorithmes évolutionnistes, ils sont des algorithmes d'optimisations stochastiques fondées sur les mécanismes de la sélection naturelle et de la génétique. Ils ont été initialement développés par John Holland en 1975. Un

algorithme génétique implémente une version très simplifiée et très schématique des mécanismes de l'évolution biologique.

Il est défini par quatre éléments de bases suivantes :

- ✓ Individu/chromosome/séquence : une solution potentielle du problème qui correspond à une valeur codée de la variable (ou des variables) en considération.
- ✓ Population : un ensemble de chromosomes ou de points de l'espace de recherche (donc des valeurs codées des variables).
- ✓ Environnement : l'espace de recherche (caractérisé en termes de performance correspondant à chaque individu possible).
- ✓ Fonction de performance (fitness) : la fonction - positive - que nous cherchons à maximiser, car elle représente l'adaptation de l'individu à son environnement.

Les AGs sont des techniques de recherche stochastique qui ont été appliqués avec succès à une large gamme de problèmes réels et complexes. C'est une méthode itérative qui utilise des opérateurs de variations stochastiques sur un "pool" d'individu (la population). Chaque individu de la population représente une version encodée (solution) du problème. Au départ, cette population initiale est engendrée aléatoirement. À chaque génération/itération de l'algorithme, les solutions sont sélectionnées, rassemblées en paires et recombinaison afin de produire de nouvelles solutions qui remplaceront les "moins bonnes" selon un certain critère, et ainsi de suite. Une fonction d'évaluation associe une valeur d'adaptation (fitness) à chaque individu dans le but de déterminer sa pertinence par rapport au problème (critère de sélection). Le principe de notre est de maximiser le score d'évaluation en fonction des valeurs de poids

En permettant aux seuls individus bien adaptés à l'environnement de se reproduire, la nature assure la pérennité de leurs meilleures caractéristiques, lesquelles caractéristiques se recombinent entre elles (chaque enfant reçoit de bonnes caractéristiques à la fois de son père et de sa mère) pour former à chaque génération de nouveaux individus toujours mieux adaptés à leur environnement.

3.4.2. Description de la solution proposée

Vu la pertinence du problème et la contribution d'une telle étude dans la résolution des difficultés précitées, nous avons essayé de traiter du sujet : « *la qualité de référencement d'articles scientifiques* ».

Dans la solution proposée, nous allons d'abord, regrouper les facteurs bibliométriques en trois groupes (auteur, revue, article).

Chaque facteur sera normalisé sur l'intervalle [0, 1]. Selon l'équation suivante :

$$\text{Équation 11 } nor = \frac{\text{valeur}_{index} - \text{valeur}_{min_{index}}}{\text{max}_{index}}$$

Une fois les facteurs normalisés, nous allons calculer des valeurs initiales de poids, en calculant la corrélation des facteurs avec d'autres facteurs de même catégories et ceci en se en fonctions des formules statistiques suivants : moyenne, minimum, maximum, premier quartile, médian. La valeur de poids initiale représentera la moyenne de corrélation de l'indicateur avec les autres facteurs.

Après cette étape, nous appliquons l'algorithme génétique en considérant :

- Les poids initiaux issus de la corrélation comme population initiale. Cette population sera représentée comme un tableau de réels $\in [0, 1]$.
- La fonction d'évaluation correspondra à la moyenne pondérée du calcul du score d'évaluation.
- La fonction de croisement sera la moyenne entre les individus de la population (moyenne entre les facteurs de chaque individu).
- La fonction de mutation, se fera pour à 1 %. Ainsi, pour un individu sur 100, nous allons générer des valeurs de poids aléatoires.

Une fois l'estimation des poids finaux réalisée avec l'algorithme génétique, le score de l'évaluation d'une référence sera calculé en fonction de la moyenne pondérée. Ce score sera calculé pour chaque catégorie de la référence (article, auteur, journal).

La section suivante présente les algorithmes de notre solution.

3.4.3. Algorithmes de la solution proposée

Le processus de la solution générale, qui représente le fonctionnement de notre outil est tel que suit :

<p>Algorithme1 : <i>Évaluation de références d'un article</i></p> <p>Début :</p> <p>1- Lancer le processus d'évaluation</p> <p>2- Récupérer les Références</p> <p>3- Pour chaque catégorie d'indicateurs (article, auteur, journal)</p> <p style="padding-left: 20px;">a) Calculer les indicateurs ou les récupérer de la base de donnée</p> <p style="padding-left: 20px;">b) Normaliser les indicateurs</p> <p style="padding-left: 20px;">c) Faire la corrélation entre ces indicateurs</p> <p style="padding-left: 20px;">d) Faire l'estimation initiale des poids des indices</p> <p style="padding-left: 20px;">e) Faire l'estimation finale des poids avec l'algorithme génétique</p> <p style="padding-left: 20px;">f) Calculer la moyenne finale des indicateurs avec leurs poids</p> <p style="padding-left: 20px;">g) Calculer l'observation</p> <p>4- Retourner le résultat final d'évaluation</p> <p>Fin</p>
--

Comme nous l'avons déjà mentionné, la catégorie représente l'un des groupes d'indicateurs suivants : article, auteur et journal.

Les indicateurs de chaque catégorie sont comme suit (voir chapitre 2 pour avoir plus de détails sur le calcul de ces facteurs) :

- Article : nombre de citations ;
- Auteur : h-index, g-index, hg-index, hr-index ;
- Journal : facteur d'impact, eigen facteur, article influence, SNIP.

Toutefois, les indicateurs utilisés peuvent très bien être étendus avec d'autres facteurs bibliométriques sans affecter la solution proposée.

3.4.3.1. La normalisation

L'algorithme suivant représente le calcul de la normalisation de chaque indicateur bibliométrique sur l'intervalle [0, 1].

Algorithme 2 : Algorithme de normalisation

Entrées : A : liste des indices;

Sorties : AN : liste des indices normalisés ;

Début :

Variables :

$n \leftarrow$ Longueur d'A

Pour i allant de 1 à n

$j \leftarrow A[j]$

$min_j \leftarrow$ minimum des valeurs du facteur j

$max_j \leftarrow$ maximum des valeurs du facteur j

$AN[i] \leftarrow (j - min_j) / max_j$

Fin pour

Retourner AN

Fin

3.4.3.2. La corrélation

Les algorithmes 3, 4, 5, 6 permettent de calculer respectivement la moyenne, la médiane, les quartiles et le minimum - maximum d'un certain indicateur.

Ces valeurs seront utilisées dans l'algorithme de corrélation (algorithme 7). La corrélation sera calculée entre les différents indicateurs d'une référence en fonction de ces valeurs (la moyenne, la médiane, les quartiles, le minimum, maximum). Ceci permettra de montrer la relation entre les différents indicateurs d'une référence au sein d'une même catégorie (article, auteur, journal). La formule de calcul de la corrélation entre deux indices X, Y est : que suit :

$$\text{Équation 12 } cor = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2} \cdot \sqrt{\sum(y-\bar{y})^2}}$$

Algorithme 3 : Moyenne des valeurs d'un indice

Entrées : A : liste des valeurs d'un indice ;

Sorties : m : moyenne ;

Début

Variables :

$S \leftarrow 0$

$n \leftarrow$ longueur de A

Pour i allant de 1 à n

$S \leftarrow S + A[i]$

Fin pour

$m \leftarrow S / n$

Retourner m

Fin

Algorithme 4 : Médiane des valeurs d'un indice

Entrées : A : liste des valeurs d'un indice;

Sorties : m : médiane des valeurs de A ;

Début

$n \leftarrow$ longueur de A ;

$B \leftarrow$ suite A rangée dans l'ordre croissant ;

Si n est impair alors

$m \leftarrow 2 \frac{(n-1)}{2} + 1$;

Sinon

$m \leftarrow \frac{1}{2} \left(2 \binom{n}{2} + 2 \binom{n}{2} + 1 \right)$;

Fin

Retourner m

Fin

Algorithme 5 : Calcul des quartiles (1^{er} et 3^{eme}) des valeurs d'un indice

Entrées : A : liste des valeurs d'un indice;

Sorties : $Q1, Q3$: premier et troisième quartiles d' A ;

Début

$n \leftarrow$ longueur de A ;

$B \leftarrow$ suite A rangé dans l'ordre croissant ;

$r \leftarrow n \% 4$ //reste de la division euclidienne de n par 4 ;

$q \leftarrow n / 4$ //quotient de la division euclidienne de n par 4 ;

Si $r = 0$ alors

$Q1 \leftarrow q$;

$Q3 \leftarrow 3q$;

Fin si

Sinon si $r = 1$ alors

$Q1 \leftarrow q+1$;

$Q3 \leftarrow 3q+1$;

Fin si

Sinon Si $r = 2$ alors

```

        Q1 ← q+1 ;
        Q3 ← 3q+2 ;
    Fin si
    Sinon Si r = 4 alors
        Q1 ← q+1 ;
        Q3 ← 3q+3 ;
    Fin si
    Retourner Q1, Q3 ;
Fin

```

Algorithme 6 : Minimum et maximum des valeurs d'un indice

Entrées : A : liste des valeurs d'un indice;

Sorties : min, max : minimum et maximum des valeurs

Début

```

    min = T[1];
    max = T[1];
    Pour i = 2 à n Faire:
        Si (T[i] < min) Alors
            min = T[i];
        Sinon Si (T[i] > max) Alors
            max = T[i];
        Fin Si
    Fin Pour
    Retourner min, max ;

```

Fin

Algorithme 7 : Corrélation entre les indicateurs d'une catégorie

```

Entrées : A : matrice des valeurs de moyenne, minimum, maximum, médiane et quartile
des indices d'une catégorie ;
Sorties : R : matrices de corrélation.
Début
    n ← longueur de A ;
    Pour i = 1 à n Faire:
        Pour j = 1 à n Faire:
            R[i, j] = cor(A[i, i], A[i, j]) //cor :
//coefficient de corrélation entre l'indice i et j

//calculé selon l'équation 11,

//cor ∈ [0, 1]

        Fin Pour
    Fin Pour
    Retourner R ;
Fin

```

3.4.3.3. Estimation des poids initiaux

L'estimation initiale des poids se fait en fonction de la matrice de corrélation *poids_init*. Pour cela, on récupère les valeurs de corrélation entre l'indice i avec tous les autres indices, puis on calcule la moyenne des ces valeurs pour avoir une estimation de l'importance de l'indice i.

A partir de cette valeur, on génère un ensemble de valeurs (selon la taille de la population de l'algorithme génétique) dans l'intervalle [*poids_init* - 0.1 , *poids_init* + 0.1].

le total des valeurs générés pour les indices représentera la population initial de l'algorithme génétique qui sera présenté dans la section suivante.

Algorithme 8 : Estimation des poids initiaux

```

Entrées : A : matrice de corrélation des indices ; // A[1..n, 1..n] avec n : nombre
d'indices
    t : taille de la population de l'algorithme génétique
Sorties : P : matrices des poids initiaux. // P[1..t, 1..n]
Début
    n ← longueur de A ;
    moy ← tableau de n éléments
    Pour i = 1 à n Faire:
        moy[i] = 0 ;
        Pour j = 1 à n Faire:
            moy[i] += A[i, j]

```

```

        Fin Pour
        moy[i] = moy[i] / n ;
    Fin Pour
    Pour i = 1 à t Faire:
        Pour j = 1 à n Faire:
            P[i, j] = random(moy[j]-0.1 , moy[j]+0.1
) ;
                //random : génère un nombre entre
(moy[i]-1 et moy[i]+1)
        Fin Pour
    Fin Pour
    Retourner P ;
Fin

```

3.4.3.4. Estimation des poids finaux

L'algorithme suivant consiste à réaliser l'estimation des valeurs de poids finaux qui seront associés aux indicateurs bibliométriques pour le calcul de la moyenne pondérée.

Ces poids sont estimés en fonction de l'algorithme génétique en prenant en compte les paramètres suivants :

- ✓ **Gène** : la valeur d'un poids W associé à un indices (un gène sera codé comme valeur réel $\in [0, 1]$).
- ✓ **Individu** : la liste des poids associés aux indicateurs (un individu est un tableau de gènes).
- ✓ **Population** : représente l'ensemble des liste des poids associés aux facteurs, ces valeurs sont calculé en fonction des corrélations entre les facteurs avec l'algorithme 8.
- ✓ **Évaluation** : la fonction d'évaluation de l'algorithme génétique sera la fonction de calcule du score d'évaluation c'est à dire la somme pondérée des indicateurs avec leurs poids : $\sum W_i F_i$.
- ✓ **Sélection** : la sélection se fera en fonction des meilleurs valeurs des poids « meilleur individus » dans notre cas on a prix le maximum ($\sum W_i F_i$).
- ✓ **Croisement** : on prend deux des meilleures solutions sélectionnées et on réalise la moyenne entre les poids des deux individus.
- ✓ **Mutation** : on modifie une des valeurs de certaines suites de manière aléatoire $\in [0,1]$ avec une probabilité de 1%.
- ✓ **Critères d'arrêt** : notre algorithme génétique peut être arrêté s'il n'y a pas d'amélioration de la fonction objective (après un certain nombre d'itérations) ou si le meilleur individu reçoit le résultat attendu.

Algorithme 9 : Estimation des poids finaux

Entrées : P : population initiale;

Sorties : poids : valeurs finales des poids.

Début

$n \leftarrow$ longueur de P . //le nombre d'individu

Pour $i = 0$ à $nb\text{-essai-max}$ **Faire:**

Pour $j = 0$ à n {

 Évaluer ($P[i]$)

Fin Pour

 Si le meilleur individu == 1 {

 Arrêt //S'il n'y a pas d'amélioration de la fonction//

Fin Pour

 Sélectionner (P)

 Croiser (P)

 Muter (P)

Fin Pour

Fin

3.4.3.5. Calcul du score d'évaluation

Enfin, l'algorithme suivant permet de calculer le score d'évaluation d'une référence. Ce score est calculé avec la moyenne pondérée en fonction des indices bibliométriques et les poids estimés à partir de l'algorithme génétique.

Algorithme 10 : Calcul du score d'évaluation

Entrées : F : tableau des indices ;

P : tableau de poids ;

Sorties : S : score d'évaluation;

Début

$s1 \leftarrow 0$;

$s2 \leftarrow 0$;

$n \leftarrow$ longueur de F ;

Pour i allant de 1 à I

$s1 = s1 + (F[i] * P[i])$;

$s2 = s2 + P[i]$;

Fin pour

$S = s1/s2$

 Retourner S

Fin

3.5. Architecture de la solution proposée

Une fois l'approche et les algorithmes de conception de notre solution présentée, nous allons étudier dans cette partie l'architecture générale de notre application. L'architecture présentée dans la figure suivante, montre le fonctionnement du système proposé pour l'évaluation des références d'un article scientifique.

Pour implémenter l'outil d'évaluation, nous avons choisi d'utiliser une architecture Client/Serveur :

- La partie cliente représentera l'interface graphique de notre application permettant à l'utilisateur de faire ses requêtes et interroger le serveur sur le score d'évaluation.
- Le serveur de données contiendra les différentes données utilisées
- Le serveur d'application, dans la première phase, récupère les références et les indices à travers le serveur de données MongoDB et calcule les indicateurs. Une fois cette étape faite, Le système passe vers la seconde phase du traitement la normalisation, corrélation, estimation et autres calculs comme le montre la figure. Une fois le score d'évaluation calculé, le serveur d'application envoie le résultat au client (l'extension Google Chrome) qui va les parser et les afficher l'utilisateur.

Le client (extension Google Chrome) en dernière étape va afficher une interface graphique incluant tous les détails sur le score d'évaluation, les indices et leurs poids, ainsi qu'une observation en fonction du score (Excellent si le score $\geq 70\%$, Bien si le score $< 70\%$ et $\geq 50\%$ et Mauvais si le score $< 50\%$).

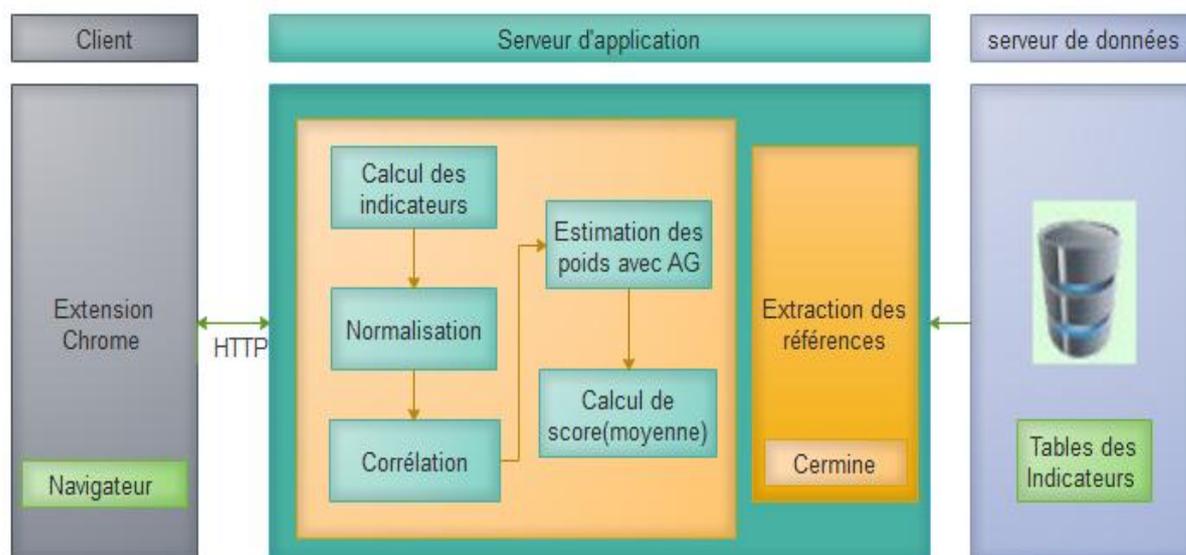


Figure 4 - Architecture générale de notre application

3.6. Modélisation de la solution

Dans cette partie nous allons présenter les différents diagrammes qui nous permettent de montrer les interactions fonctionnelles entre l'acteur et le système.

3.6.1. Diagramme de cas d'utilisation

Notre système interagit principalement avec un seul type d'utilisateur. Le diagramme de cas d'utilisation va définir les différentes actions que l'acteur pourra effectuer.

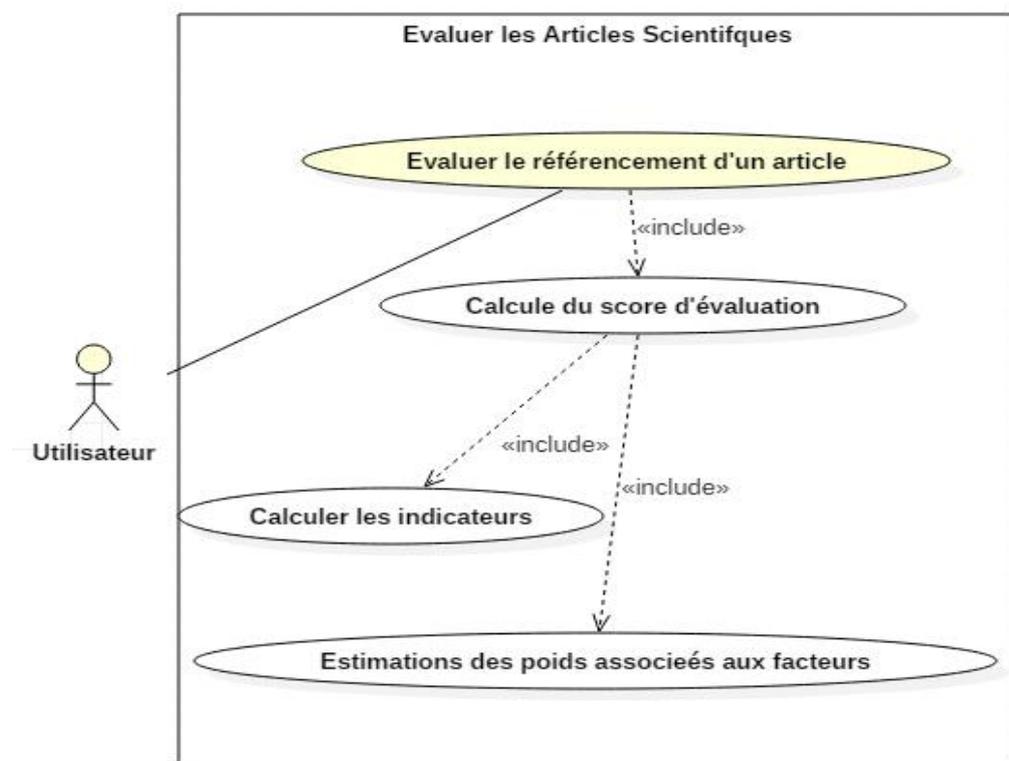


Figure 5 - Diagramme de cas d'utilisation

L'acteur est une entité en interaction avec le système. Il est l'idéalisation d'un rôle joué par une personne externe, un processus ou une chose qui interagit avec l'application. L'utilisateur du système pourra réaliser l'action Évaluer afin d'utiliser le système.

3.6.2. Diagramme de classe

Le diagramme de classe illustré dans la figure suivante représente la partie du système qui va persister. Dans la suite, chaque classe va être décrite pour une meilleure compréhension du diagramme :

Références : Elle contient les informations sur les références.

Journal : Classe héritant d'Indice-Journal contient les informations sur le journal.

Auteur : Classe héritant d'Indice-Auteur décrivant un médicament avec un nom, prénom et toutes ses coordonnées.

Publication : Classe héritant d'Indice-Publication. Elle décrit les différentes informations sur la publication scientifique l'auteur, le titre le nom du journal etc...

Les trois classes des Indices (auteur, journal, publication) sont des classes héritant de la classe Indicateurs.

Index-auteur : Cette classe regroupe l'ensemble des indicateurs de la classe Auteur.

Index-journal : Regroupe l'ensemble des indicateurs de la classe Journal.

Index-publication : Cette classe regroupe l'ensemble des indicateurs de la classe Publication

Indicateurs : cette classe contient les informations sur les indicateurs et les différentes opérations qu'elle va faire.

Evaluation : contient deux attributs sur l'évaluation et fait le calcul du score d'évaluation et l'affichage de l'observation.

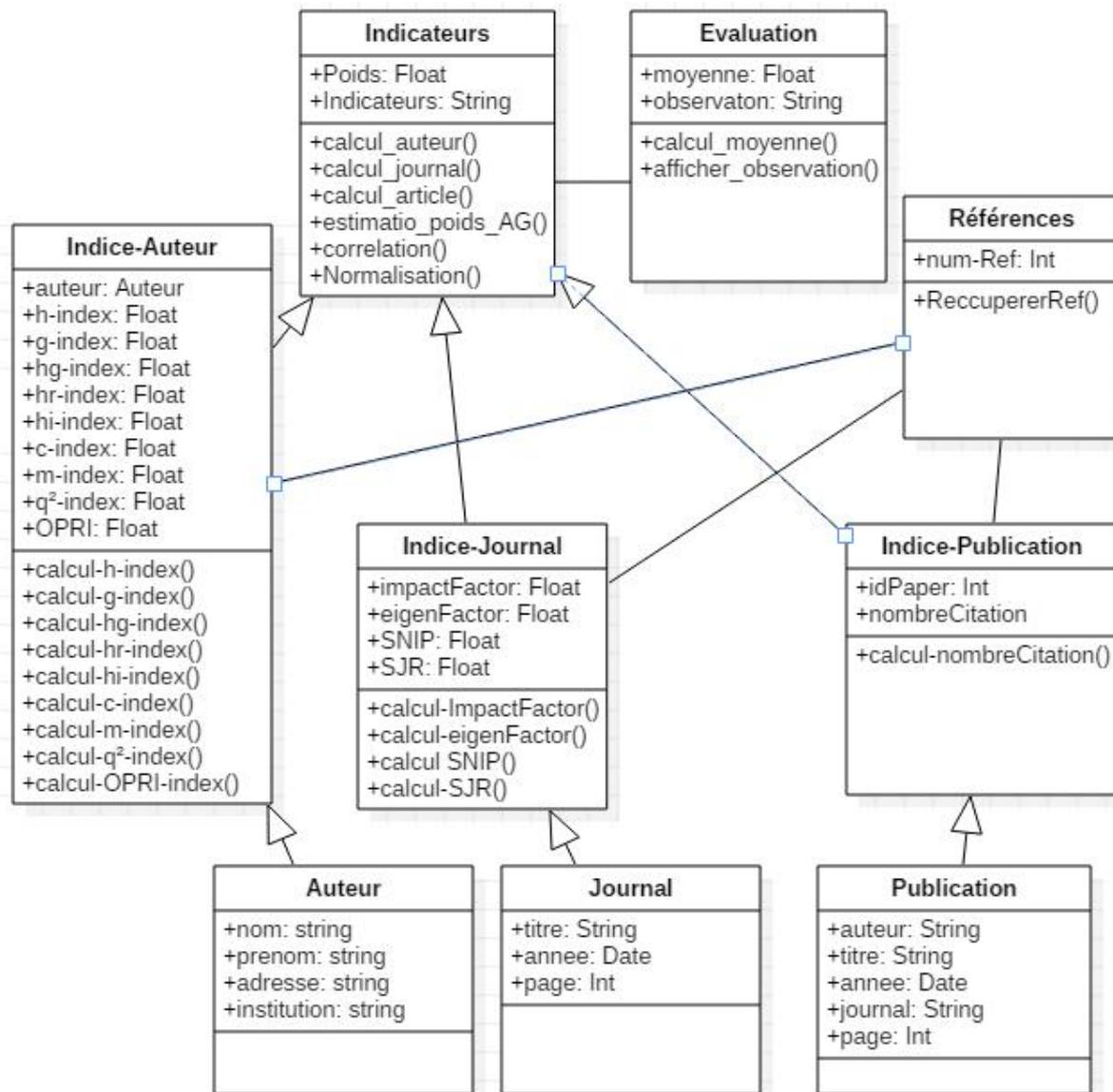


Figure 6 - Diagramme de classe

3.6.3. Diagramme d'activité

Le diagramme d'activité illustré dans cette figure Représente les règles d'enchaînement des activités et actions dans le système il y a deux parties :

- 1) Partie client : L'utilisateur ouvre un article scientifique, il clique sur le bouton représentant l'outil développé.
- 2) Partie serveur : Le serveur fait son traitement :
 - Télécharger l'article scientifique ;
 - Extraire les références ;
 - Calculer et récupérer les indicateurs ;
 - Normaliser les indicateurs ;
 - Corréler les indicateurs entre eux ;

- Estimer les poids initiaux ;
- Estimer les poids finaux avec l'algorithme génétique ;
- Calculer le score et l'observation.

Et là on revient à la partie client qui montre l'affichage de l'interface avec les résultats.

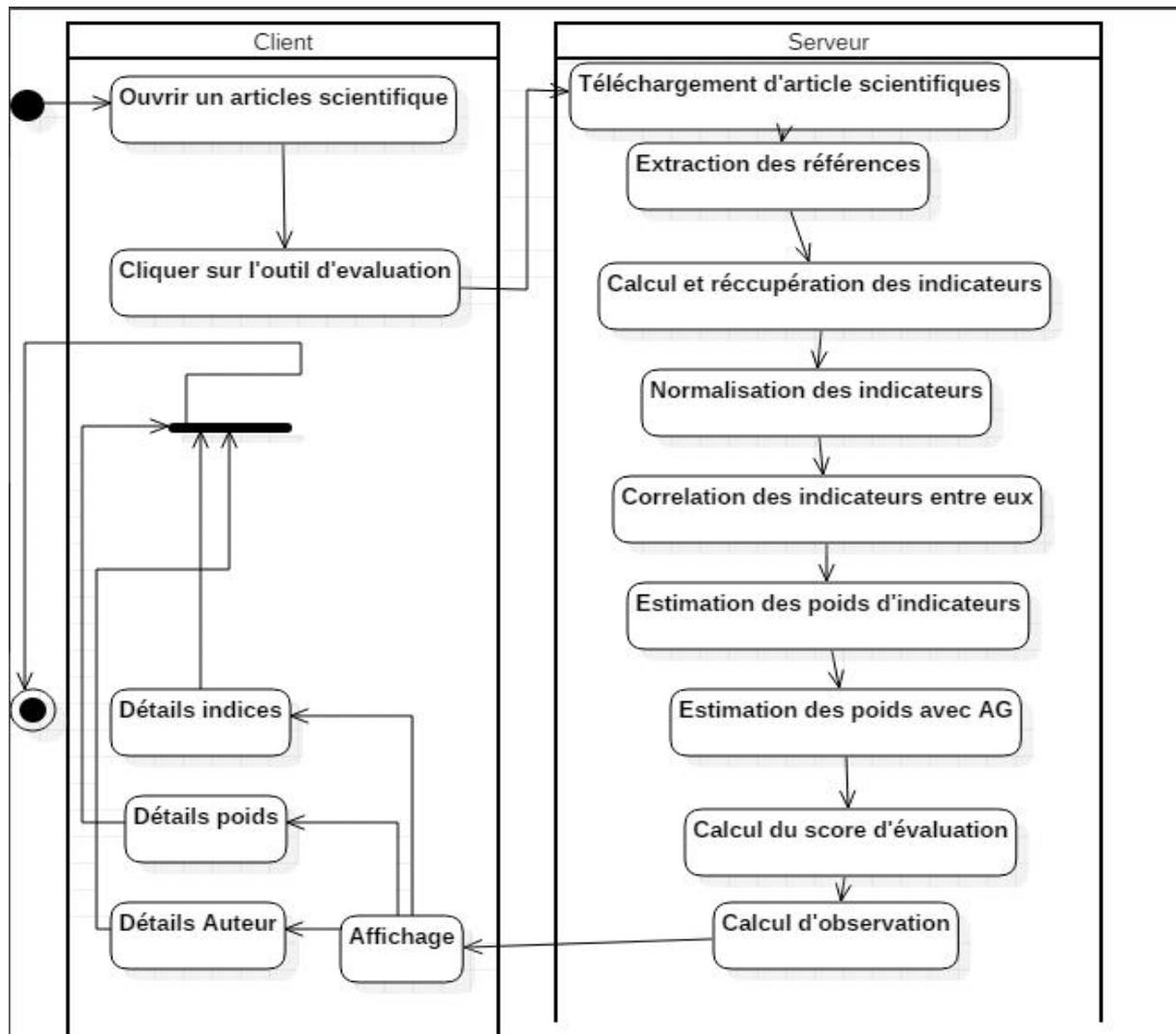


Figure 7 - Diagramme d'activité

3.6.4. Diagramme de séquence

Le diagramme de séquence illustré dans la figure 8 se déroule comme suit :

- L'utilisateur utilise l'extension Google chrome pour évaluer les références d'un article scientifique.
- Une fois L'extension lancée, le système récupère les références du serveur de données ou les calcule.
- Le serveur de donnée donne le résultat au serveur d'application.

- Le serveur d'application fait les traitements et envoie les résultats sous forme JSON à l'extension.
- L'extension analyse le résultat et l'envoie au utilisateur sous forme HTML.

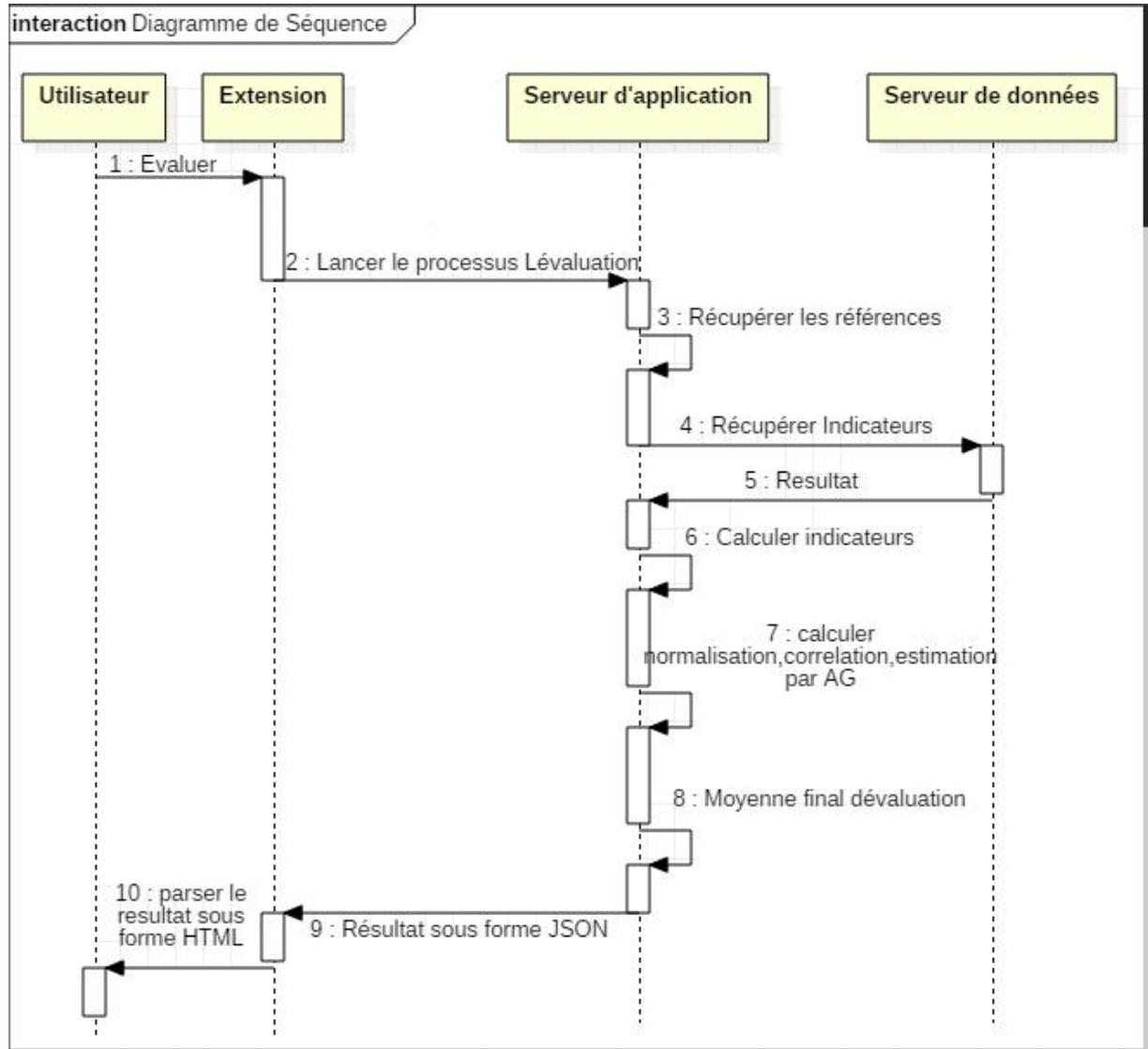


Figure 8 - Diagramme de Séquence

3.7. Conclusion

Nous avons présenté à travers ce chapitre la solution proposée et les approches de conception suivies. Nous avons aussi détaillé la modélisation de notre système en utilisant UML, nous avons établi les diagrammes nécessaires à la modélisation qui respecte bien les normes du langage UML. Enfin, nous avons défini, le schéma qui présente l'architecture de notre application.

Dans le chapitre suivant, nous allons présenter l'implémentation et l'utilisation de notre outil.

Chapitre 4 :

Implémentation et Résultat

4.1. Introduction

Dans le chapitre précédent nous avons vu la partie conception de notre solution Dans le présent chapitre, nous allons dans un premier temps présenter les différents logiciels utilisé pour le développement et l'implémentation de l'outil proposé. Dans un deuxième temps, nous allons exposer quelques exemples de cas d'utilisation de l'outil d'évaluation.

4.2. Langages de programmation

Les langages de programmation utilisés pour la réalisation de notre solution sont :

4.2.1. JavaScript

Le JavaScript est un langage de programmation de scripts orienté objet. Un script est un programme dont le code source est inclus dans un document HTML. Ce programme est interprété et s'exécute sur la machine du client lorsque le document est chargé ou lors d'une action de l'utilisateur (clic ou déplacement du curseur par exemple).

Les principes, l'implémentation, le fonctionnement, les domaines d'application de JavaScript n'ont strictement rien à voir avec ceux de Java. Seule une éventuelle ressemblance superficielle au niveau de la syntaxe, et la volonté de profiter d'un effet de mode relatif à Java ont valu à ce langage de s'appeler JavaScript [19]

4.2.2. HTML

Le HTML (Hyper Text Markup Langage) est un langage universel utilisé sur le World Wide Web Ce langage est utilisé pour créer des pages web. L'acronyme signifie *HyperText Markup Language*, ce qui signifie en français "*langage de balisage d'hypertexte*". Cette signification porte bien son nom puisqu'effectivement ce langage permet de réaliser de l'hypertexte à base d'une structure de balisage[19]. Le HTML permet de :

- Publier des documents sur le web.
- Retrouver des informations en un click de souris grâce aux Liens Hypertexte.
- Concevoir des formulaires permettant d'envoyer et de recevoir des infos sur le Web.

- Insérer directement des documents dans d'autres formats, des sources vidéo et sonores et d'autres applications.

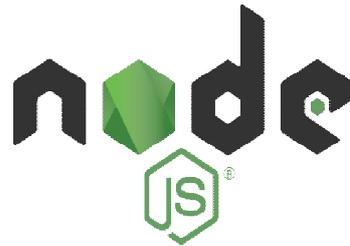
4.2.3. CSS

Le terme **CSS** est l'acronyme anglais de *Cascading Style Sheets* qui peut se traduire par "feuilles de style en cascade". Le CSS est un langage informatique utilisé sur l'internet pour mettre en forme les fichiers HTML ou XML. Ainsi, les feuilles de style, aussi appelé les fichiers CSS, comprennent du code qui permet de gérer le design d'une page en HTML. [19]

4.3. Environnement et outils de développement

Les environnements et outils utilisés pour l'implémentation de notre outil sont tels que suit :

4.3.1. Node.js



Node.js est une plateforme de développement JavaScript. Ce n'est pas vraiment un serveur, ou une Framework, c'est juste le langage JavaScript avec des bibliothèques permettant de réaliser des actions comme écrire sur la sortie standard, ouvrir/fermer des connections réseau ou encore créer un fichier. [20]

4.3.2. Mongo dB et Mongoose



MongoDB est un SGBD NoSQL open source orienté documents. Il prend en charge des types de données très divers. Le modèle de données documentaire de MongoDB permet aux développeurs de stocker et de combiner des données, quelle que soit leurs structures, sans sacrifier l'accès à ces données ni leur indexation. Ainsi, les administrateurs de base de donnée peuvent modifier dynamiquement le schéma sans aucune interruption de service. [21]

mongoose

Mongoose est une bibliothèque ODM (Object Data Modeling) pour MongoDB et Node.js. Il gère les relations entre les données, fournit la validation du schéma et est utilisé pour la conversion entre les objets du code et la représentation de ces objets dans MongoDB. [22]

4.4.3. Cermine

CERMINE est un système open source complet permettant d'extraire des métadonnées et du contenu d'articles scientifiques sous forme numérique d'origine. Le système est capable de traiter des documents au format PDF et d'extraire : les métadonnées du document, y compris le titre, les auteurs, les affiliations, les résumés, les mots-clés, le nom de la revue, le volume et le numéro, références bibliographiques analysées la structure des sections du document, les titres des sections et les paragraphes.

CERMINE repose sur un flux de travail modulaire, dont l'architecture garantit que chaque étape du flux de travail peut être gérée séparément. En conséquence, il est facile d'effectuer une évaluation, une formation, d'améliorer ou de remplacer une implémentation en une étape sans changer les autres parties du flux de travail. La plupart des étapes utilisent des techniques d'apprentissage supervisées et non supervisées, ce qui augmente la facilité de maintenance du système, ainsi que sa capacité à s'adapter aux nouvelles présentations de documents. [23]

4.4. Architecture d'implémentation du système

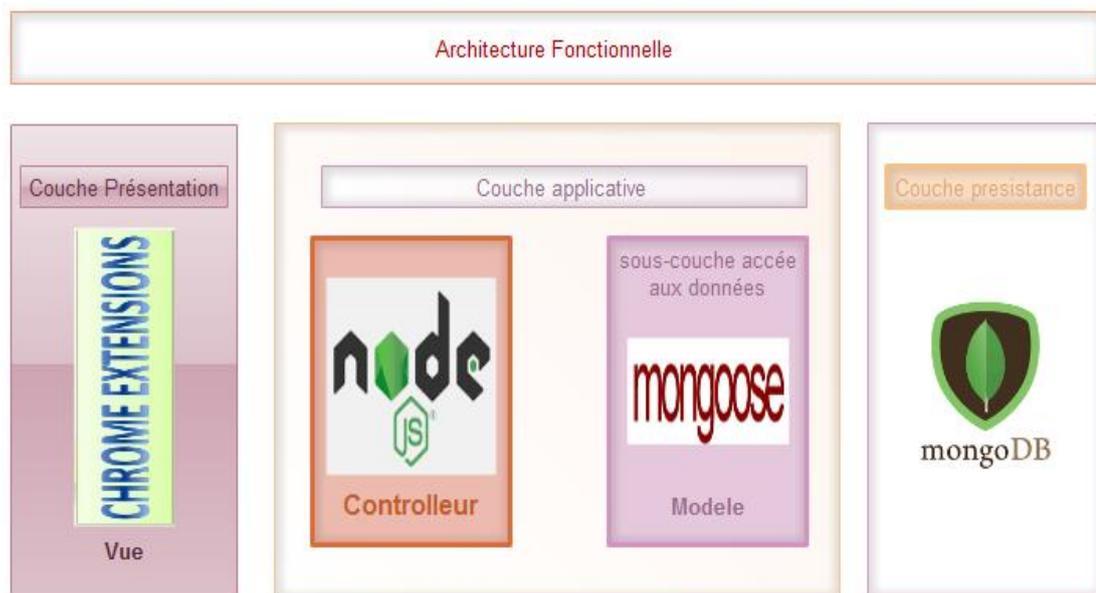


Figure 9 - Architecture fonctionnelle

Notre système est composé de trois principales couches, qui elles-mêmes sont éventuellement composées de deux sous couches comme illustré dans la Figure 4.1.

Les trois principales couches sont les suivantes :

- ✓ La couche persistance des données : Elle contient notre SGBD NoSQL et s'occupe de la gestion des données.
- ✓ La couche applicative : contient le service web et la logique métiers.
- ✓ La couche présentation : c'est l'interface graphique sur le web.

L'architecture fonctionnelle de notre système est basée sur le modèle de conception MVC. Le but de cette architecture est de séparer les aspects traitement, présentation et persistance. Elle a aussi l'avantage d'offrir un faible couplage entre les composantes de l'architecture, ce qui apporte une flexibilité pour une future maintenance du système ou le remplacement d'une technologie devenue obsolète. Les données sont gérées par le modèle, la présentation par la vue, les traitements par des actions et l'ensemble est coordonné par les contrôleurs. Les fonctions des différentes couches sont :

- **La couche persistance** : pour le stockage des données, la modélisation objet proposée est transformée en des documents (format JSON), où chaque classe est soit un document ou un attribut composé à l'intérieur d'un document.
- **La couche applicative** : cette couche est responsable des traitements des requêtes des clients et de la récupération des réponses.
- **La couche présentation** : la couche présentation est matérialisée par l'IHM et c'est à travers elle que l'utilisateur exploite le côté serveur. Elle dialogue avec le serveur pour l'envoi et la réception de messages qui sont échangés avec le serveur sous la forme JSON et envoyés avec le protocole HTTP.

4.5. Cas d'utilisations

4.5.1. Lancer le serveur de données mongoDB

Comme première étape on commence par la commande **mongod** pour lancer notre serveur mongod.

```

osboxes@osboxes:~$ mongod --storageEngine=mmapv1 --dbpath /home/osboxes/evaluation/data
2019-06-11T14:41:33.138-0400 I CONTROL [main]
2019-06-11T14:41:33.143-0400 W CONTROL [main] 32-bit servers don't have journaling enabled by default. Please use --journal if you want durability.
2019-06-11T14:41:33.144-0400 I CONTROL [main]
2019-06-11T14:41:33.152-0400 I CONTROL [initandlisten] MongoDB starting : pid=1461 port=27017 dbpath=/home/osboxes/evaluation/data 32-bit host=osboxes
2019-06-11T14:41:33.152-0400 I CONTROL [initandlisten] db version v3.2.22
2019-06-11T14:41:33.152-0400 I CONTROL [initandlisten] git version: 105acca0d443f9a47c1a5bd608fd7133840a58dd
2019-06-11T14:41:33.153-0400 I CONTROL [initandlisten] allocator: tcmalloc
2019-06-11T14:41:33.153-0400 I CONTROL [initandlisten] modules: none
2019-06-11T14:41:33.153-0400 I CONTROL [initandlisten] build environment:
2019-06-11T14:41:33.153-0400 I CONTROL [initandlisten]     distarch: i686
2019-06-11T14:41:33.153-0400 I CONTROL [initandlisten]     target_arch: i386
2019-06-11T14:41:33.153-0400 I CONTROL [initandlisten] options: { storage: { db
Path: "/home/osboxes/evaluation/data", engine: "mmapv1"
} }
2019-06-11T14:41:33.261-0400 I CONTROL [initandlisten]
2019-06-11T14:41:33.261-0400 I CONTROL [initandlisten] ** WARNING: This 32-bit MongoDB binary is deprecated
2019-06-11T14:41:33.261-0400 I CONTROL [initandlisten]
2019-06-11T14:41:33.261-0400 I CONTROL [initandlisten]
2019-06-11T14:41:33.261-0400 I CONTROL [initandlisten] ** NOTE: This is a 32-bit

```

```

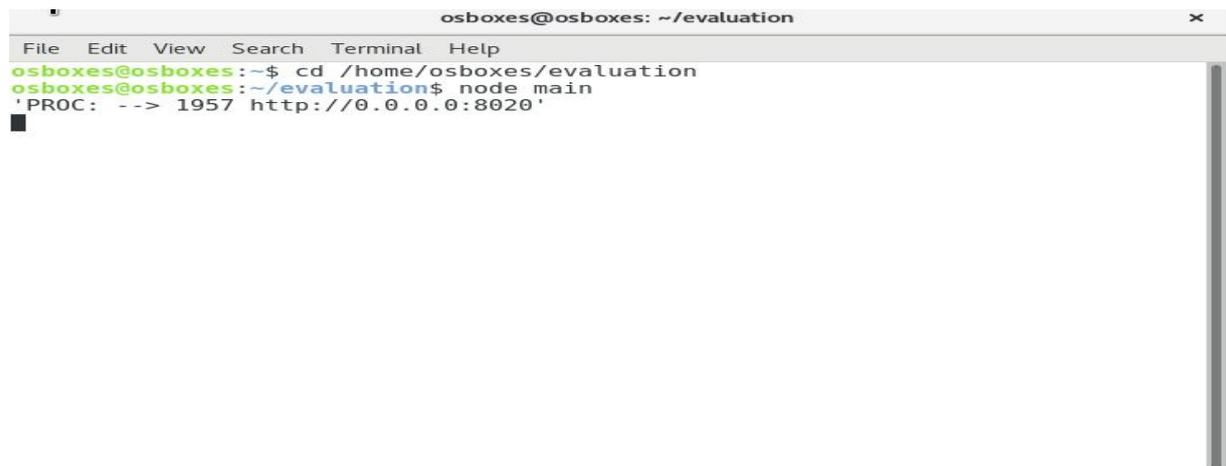
osboxes@osboxes: ~
File Edit View Search Terminal Help
mongod(_ZN5mongo15printStackTraceERSo+0x33) [0x8f471f3]
mongod(+0xEFDF60) [0x8f45f60]
mongod(+0xEFE830) [0x8f46830]
linux-gate.so.1(__kernel_sigreturn+0x0) [0xb76ffd04]
linux-gate.so.1(__kernel_vsyscall+0x9) [0xb76ffc9]
libc.so.6(gsignal+0xB0) [0xb74b7dd0]
libc.so.6(abort+0x157) [0xb74b9297]
mongod(_ZN5mongo15invariantFailedEPKcS1_j+0xE0) [0x8ecc930]
mongod(_ZN5mongo14NamespaceIndex4initEPNS_16operationContextE+0x6C8) [0x8c310d8]
]
mongod(_ZN5mongo26MMAPV1DatabaseCatalogEntryC1EPNS_16operationContextENS_10StringDataE53_bb+0x1BB) [0x8c72a2b]
mongod(_ZN5mongo12MMAPV1Engine23getDatabaseCatalogEntryEPNS_16operationContextENS_10StringDataE+0x16C) [0x8c7740c]
mongod(_ZN5mongo14DatabaseHolder6openDbEPNS_16operationContextENS_10StringDataEPb+0x3F7) [0x868e217]
mongod(+0x4CBC21) [0x8513c21]
mongod(+0x4CF032) [0x8517032]
mongod(main+0x333) [0x851bb43]
libc.so.6(__libc_start_main+0xF6) [0xb74a4286]
mongod(+0x4CA681) [0x8512681]
----- END BACKTRACE -----
Aborted
osboxes@osboxes:~$

```

Figure 10 - Lancement de mongoDB

4.5.2. Lancement du serveur de calcul

Après le lancement de la base de données, nous démarrons le serveur node.js comme le montre la figure suivante :

A terminal window titled 'osboxes@osboxes: ~/evaluation' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal shows the following commands and output:

```
osboxes@osboxes:~$ cd /home/osboxes/evaluation
osboxes@osboxes:~/evaluation$ node main
'PROC: --> 1957 http://0.0.0.0:8020'
```

Figure 11 - Lancement du serveur nodejs

4.5.3. Lancement de l'évaluation

Maintenant, nous pouvons tester notre solution. Pour cela, nous ouvrons un article scientifique en ayant installer au préalable l'extension chrome représentant la partie cliente de notre solution, et on clique sur le bouton pour lancer l'évaluation.

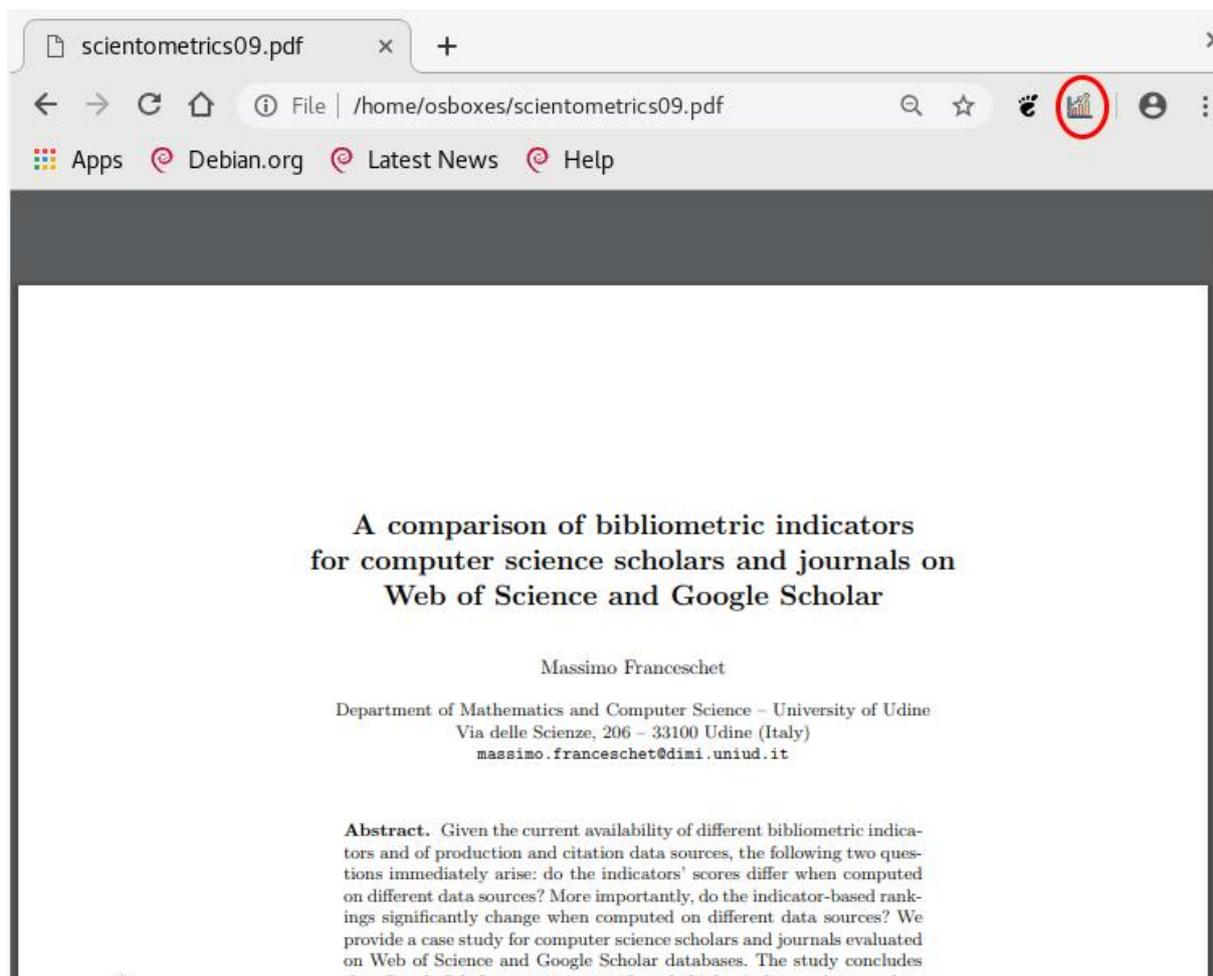
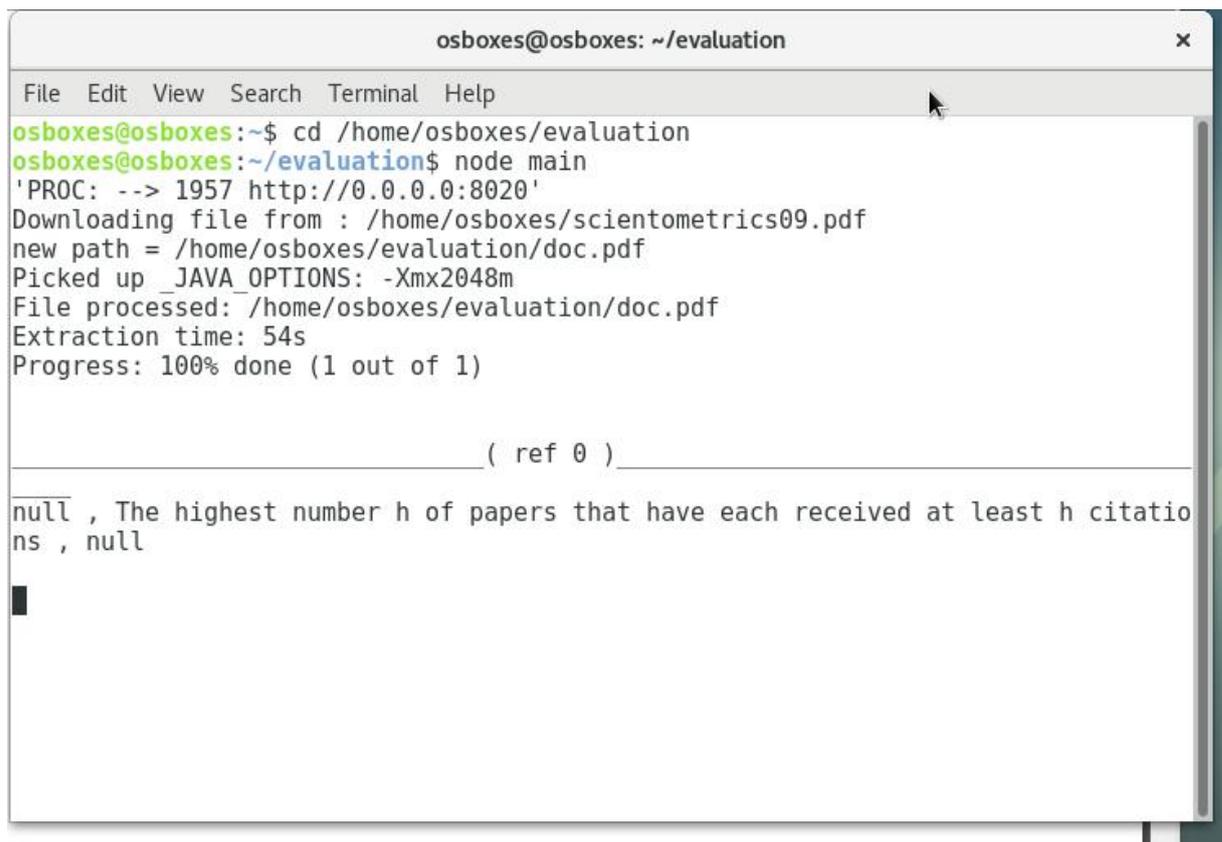


Figure 12 - Lancement de l'évaluation d'un article

Une fois qu'on clique sur le bouton de l'extension  le serveur **nodejs** commence la récupération des références, le calcul des indicateurs et des poids et le renvoi des résultats.

A terminal window titled "osboxes@osboxes: ~/evaluation" with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal shows the following output:

```
osboxes@osboxes:~$ cd /home/osboxes/evaluation
osboxes@osboxes:~/evaluation$ node main
'PROC: --> 1957 http://0.0.0.0:8020'
Downloading file from : /home/osboxes/scientometrics09.pdf
new path = /home/osboxes/evaluation/doc.pdf
Picked up _JAVA_OPTIONS: -Xmx2048m
File processed: /home/osboxes/evaluation/doc.pdf
Extraction time: 54s
Progress: 100% done (1 out of 1)

_____ ( ref 0 ) _____
null , The highest number h of papers that have each received at least h citations , null
█
```

Figure 13 - Traitements en cours du serveur

Figure

(ref 6)

N. Bakkalbasi , Three options for citation tracking: Google Scholar, Scopus and Web of Science , Biomedical Digital Libraries

poids_article_numcit : 1

Num Citation: 25

Num Citation nor: 0.05

Moyenne ARTICLE : 0.05

observation ARTICLE : Mauvais

Indice H : 4

Indice G : 18

Indice G : 8.48528137423857

Indice H nor: 0.08

Indice G nor: 0.36

Indice G nor: 0.1697056274847714

[0.3300078096395122, 0.3216010325263832, 0.3483911578341046]

poids_author_hindex : 0.3300078096395122

poids_author_gindex : 0.3216010325263832

poids_author_gindex : 0.3483911578341046

Moyenne AUTHOR : 0.0671003121770139

observation AUTHOR : Mauvais

Revue = Biomedical Digital Libraries

Journal : Biomedical Digital Libraries

factors not found...

(ref 7)

Bar-Ilan , Which h-index? A comparison of WoS, Scopus and Google Scholar , Scientometrics

14 -

4.5.4. Affichage des résultats finaux

Une fois que le serveur a terminé son calcul, il renvoie les résultats finaux à l'extension chrome qui va les afficher sous forme de page HTML.

Cette interface permet d'afficher l'ensemble des résultats finaux à l'utilisateur : elle affiche entre autres les valeurs des indices bibliométriques avec leurs poids, les détails d'auteurs ainsi que la moyenne et l'observation final

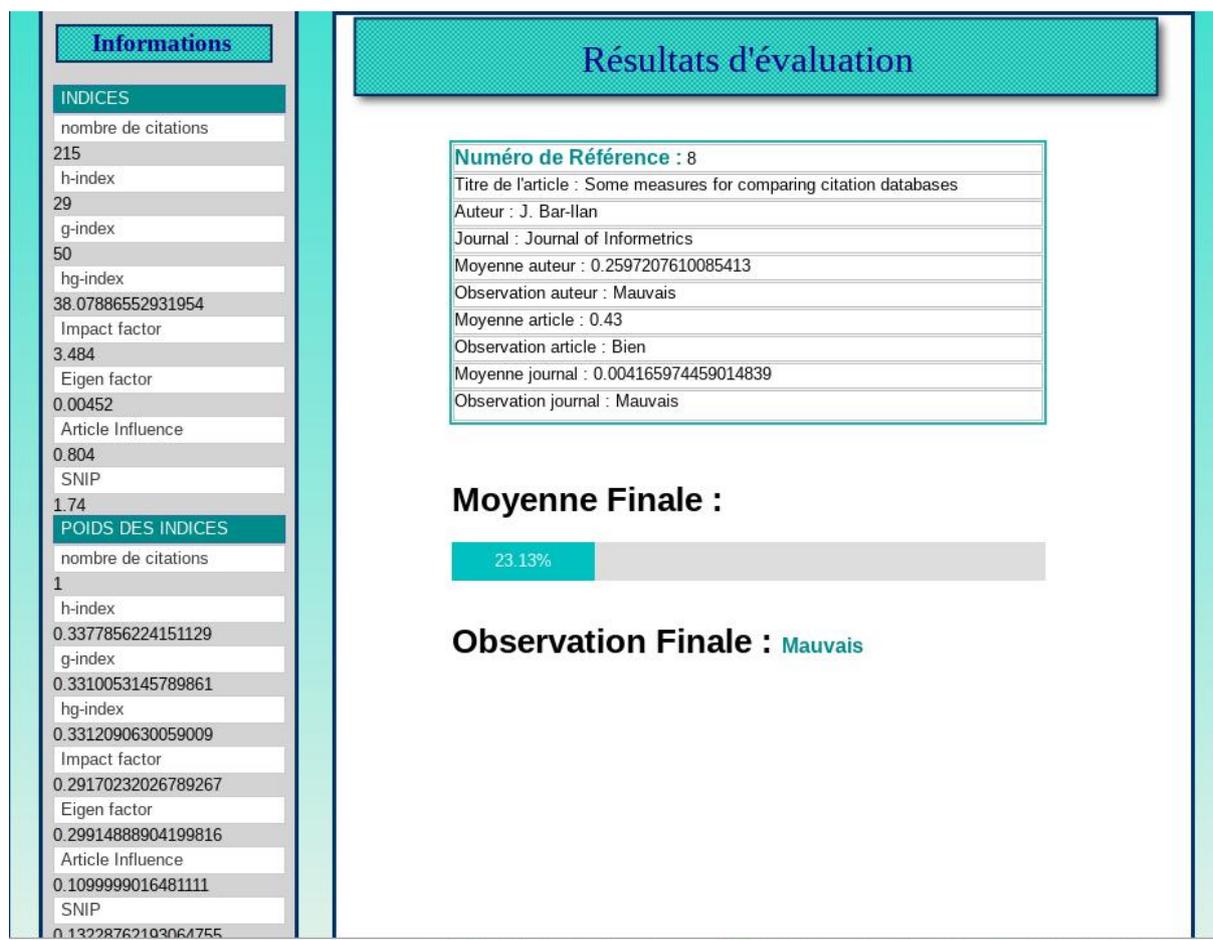


Figure 15 - Interface des résultats finaux

4.6. Conclusion

Nous avons vu dans ce chapitre les différentes étapes d'implémentation et de test de l'outil proposé.

Ainsi, nous avons pu finalisé la dernière étape de l'expérimentation et le résultat de ce projet de fin d'études.

Conclusion général

Dans ce travail, nous nous sommes intéressés à la recherche scientifique afin de connaître l'évaluation des références bibliographiques d'un article.

Ce projet de fin d'étude a consisté à concevoir un outil permettant de réaliser cette évaluation. Pour concevoir ce travail nous avons présenté premièrement quelques définitions sur la recherche scientifique d'une façon générale et l'évaluation de la recherche scientifique, nous avons aussi présenté la bibliométrie et les différents indicateurs bibliométriques.

Notre étude scientifique a été dirigée dans le deuxième chapitre vers les méthodes et les stratégies de l'évaluation scientifiques, on a vu beaucoup d'informations sur ces stratégies, dont certaines sont encore en développement et d'autres commencent à être utilisées.

Dans le troisième chapitre nous avons développé notre solution et détaillé les étapes de conception et de modélisation de la solution proposée.

En dernier, nous avons présenté les phases d'implémentation et de réalisation de notre travail. L'outil proposé a été développé sous la forme d'une extension chrome.

Cependant, ce travail étant une œuvre humaine, n'est pas un modèle unique et parfait, c'est pourquoi nous restons ouverts à toutes les critiques et nous sommes prêts à recevoir toutes les suggestions et remarques tendant à améliorer d'avantage cette étude.

En dernier, nous pouvons conclure que l'évaluation de la recherche scientifique est une procédure mise en place permettant de contribuer au financement d'activités de recherche, pour évaluer le bon usage de son évolution. Il en existe plusieurs modalités possibles. L'évaluation peut concerner toutes sortes d'entités de la recherche scientifique, de l'évaluation individuelle à celle d'universités ou d'instituts. Notre proposition essaie de faciliter le travail aux différents chercheurs afin d'avoir une première idée d'un article

Références bibliographiques

[1] **Blandine Laperche.** « Les critères marchands d'évaluation du travail scientifique dans la nouvelle économie La science comme “force productive” et “outil marketing » ». « Revue d'économie et de management de l'innovation ». 2003. pages 105 à 138.

[2] **Pr. Marzuki B. Khalid.** « La recherche et ses méthodologies ». Disponible à l'adresse : <http://www2.ift.ulaval.ca/~chaib/IFT-6001/Slides/Rech-method.pdf>. 48 pages.

[3] **ARSOM.** Offrant des instruments pour l'évaluation de la recherche pour le développement. Disponible à l'adresse : https://www.guidelines.kaowarsom.be/fr/evaluation_articles_scientifiques.

[4] petit guide d'évaluation participative à l'intention des initiatives de développement des communautés. Disponible sur lien <http://www.rechercheparticipative.org/index.html>.

[5] **[CRDE]**Centre de Recherche et de Développement en Education. Université Moncton Canada. <http://web.umoncton.ca/umcm-longd01/TheorixDownload/module11.pdf>

[6] **Université LAVAL.** Chercher une ressource documentaire. Disponible à l'adresse : <https://www.bibl.ulaval.ca/services/bibliometrie>.

[7] **CoopIST.** Indicateurs de notoriété des publications. Disponible à l'adresse : <https://coop-ist.cirad.fr/aide-a-la-publication/evaluer-les-publications/revue/indicateurs-de-notoriete/1-impact-factor-ou-facteur-d-impact-if-ou-fi>.

[8] **Ibáñez, Alfonso,** et al. Genetic algorithms and Gaussian Bayesian networks to uncover the predictive core set of bibliometric indices. Journal of the Association for Information Science and Technology, 2015.

[9] **[GHA2016]** GHAMNIA Leyla, évaluation d'un outils de calcul de la qualité de référencement d'articles scientifiques, UNIVERSITE ABDELHAMID IBN BADIS DE MOSTAGANEM, 2016, 52 pages.

[10] **Julien Sico.SlideShare.** Disponible à l'adresse : <https://fr.slideshare.net/jsicot/google-scholar-un-moteur-de-recherche-pour-linformation-scientifique>.

[11] **Fabien Soyez.** TECHNIQUES DE L'INGÉNIEUR L'ÉDITEUR TECHNIQUE ET SCIENTIFIQUE DE RÉFÉRENCE. Disponible à l'adresse : <https://www.techniques-ingenieur.fr/actualite/articles/moteurs-de-recherche-scientifiques-les-alternatives-a-google-scholar-1621/#reagir-article.2/31/2014>

[12] **Taboola.**Le magazine cloud & big data. Disponible à l'adresse : <https://www.lebigdata.fr/machine-learning-et-big-data>

[13] **Alemdar, N. & Özyildirim, S.** (1998), 'A genetic game of trade, growth and externalities', Journal of Economic Dynamics and Control 22(6), 811–832.

[14] **Research Gate.**Application des algorithmes génétiques aux problèmes d'optimisation. Disponible à l'adresse :

<https://www.researchgate.net/publication/261732872> Application des algorithmes génétiques aux problèmes d'optimisation

[15] **Rasim M. Alguliyev et Nigar T. Ismayilova.** IMPACT FACTOR WEIGHTED BY IMPACT FACTOR .2015,14 pages.

[16] **Yoshiko Okubo.** Indicateurs bibliométriques et analyse des systèmes de recherche : Méthodes et exemples, Éditions OCDE 1997, 69 pages.

[17] **Thomas Vallée et Murat Yıldızoglu,** Présentation des algorithmes génétiques et de leurs applications en économie.

[18] **Weller, Katrin.** Social media and altmetrics. "An overview of current alternative approaches to measuring scholarly impact. Incentives and performance. Springer International Publishing, 2015. Pg. 261-276.

[19] Javascript, HTML,CSS <http://glossaire.infowebmaster.fr/html/> visité le 20/05/2019.

[20] Mathieu Nebra. DES APPLICATIONS ULTRA-RAPIDES AVEC NODE.JS.Openclassrooms.11/03/2014.

[21] Openclassrooms. Découvrez le fonctionnement de MongoDB.Disponible à l'adresse : <https://openclassrooms.com/fr/courses/4462426-maitrisez-les-bases-de-donnees-nosql/4474601-decouvrez-le-fonctionnement-de-mongodb>

[22] Jamie Munro. An Introduction to Mongoose for MongoDB and Node.js.Disponible à l'adresse: <https://code.tutsplus.com/articles/an-introduction-to-mongoose-for-mongodb-and-nodejs--cms-29527>

[23] Dominika Tkaczyk,CERMINE: automatic extraction of structured metadata from scientific literature, December 2015, Volume 18, Issue 4, pp 317–335.

