



HAL
open science

Contribution à la conception et à la maîtrise du processus de référencement Web

Sylvain Sagot

► **To cite this version:**

Sylvain Sagot. Contribution à la conception et à la maîtrise du processus de référencement Web. Autre. Université de Technologie de Belfort-Montbéliard, 2016. Français. NNT : 2016BELF0310 . tel-01492889

HAL Id: tel-01492889

<https://theses.hal.science/tel-01492889v1>

Submitted on 20 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SPIM

Thèse de Doctorat



école doctorale sciences pour l'ingénieur et microtechniques
UNIVERSITÉ DE TECHNOLOGIE BELFORT-MONTBÉLIARD

THÈSE

Présentée par

Sylvain SAGOT

Pour obtenir le grade de Docteur

Spécialité : Sciences pour l'ingénieur (Génie Informatique)

Contribution à la conception et à la maîtrise du processus de référencement Web

Soutenue publiquement le 8 Décembre 2016

Jury

Patrick BRÉZILLON

Professeur des Universités – UPMC, Paris

Rapporteur

Jean-Claude DOMENGET

Maître de Conférences – UBFC, Montbéliard

Examineur

Alain-Jérôme FOUGÈRES

Enseignant-Chercheur HDR – ESTA, Belfort

Directeur de thèse

Myriam LEWKOWICZ

Professeur des Universités – UTT, Troyes

Examinatrice

Egon OSTROSI

Maître de Conférences – UBFC, Belfort

Co-directeur de thèse

Philippe TRIGANO

Professeur des Universités – UTC, Compiègne

Rapporteur

Remerciements

Ce travail de thèse de doctorat n'aurait pu être conduit sans l'aide et le soutien d'un grand nombre de personnes. C'est pourquoi je tiens à leur adresser mes sincères remerciements :

Je témoigne ma profonde reconnaissance à mon directeur de thèse, Monsieur Alain-Jérôme Fougères, Enseignant-Chercheur (HDR) à l'École Supérieure de Technologie et des Affaires (ESTA, Belfort), pour l'accueil qu'il m'a réservé dans son équipe de recherche et pour l'aide, le soutien et le temps qu'il m'a accordés.

J'exprime ma profonde gratitude à mon co-directeur de thèse, Monsieur Egon Ostrosi, Maître de Conférences à l'Université de Technologie de Belfort-Montbéliard (Belfort, UBFC), pour ses précieux conseils, ses compétences et son expérience, pour sa participation active au cheminement de cette thèse, pour ses réponses à mes interrogations et la confiance qu'il m'a accordée.

Je témoigne ma sincère gratitude à Messieurs Fabien Guillosoy et Gonzague Drion, dirigeants des Entreprises ORIXA MEDIA/VISIPERF, pour l'accueil qu'ils m'ont réservé dans leur entreprise, pour la confiance qu'ils m'ont accordée durant tout ce parcours, et pour l'intérêt des missions qu'ils m'ont confiées.

J'exprime ma gratitude pour l'honneur qu'ils me font d'avoir accepté de juger ce travail à Messieurs les rapporteurs,
Monsieur Patrick Brézillon, Professeur des Universités à l'Université Pierre et Marie Curie (UPMC, Paris VI)
Monsieur Philippe Trigano, Professeur des Universités à l'Université de Technologie de Compiègne (UTC, Compiègne),
Ainsi qu'à Messieurs les examinateurs,
Monsieur Jean-Claude Domenget, Maître de Conférences à l'Université de Franche-Comté (UBFC),
Madame Myriam Lewkowicz, Professeur des Universités à l'Université de Technologie de Troyes (UTT).

J'adresse mes plus chaleureux remerciements à Monsieur Jean-Grenier Godard, directeur de l'École Supérieure de Technologie et des Affaires (ESTA, Belfort) pour l'accueil qu'il m'a réservé dans son école, et pour la con-

Remerciements

fiance sans faille qu'il m'a accordée tout au long de ces années d'étude et de travail.

J'adresse mes sincères remerciements à mes collègues de travail, et en particulier à Anthony, pour leur soutien et leur compréhension.

J'exprime ma sincère reconnaissance à mes amis et à tous ceux qui m'ont soutenu et encouragé.

J'adresse mes chaleureux remerciements à ma mamie pour sa gentillesse, sa présence et son attention.

Je remercie ma sœur Lucie qui a également choisi de se lancer dans de longues études ainsi que son compagnon Benoît.

J'exprime mon immense gratitude à mes parents, pour leur soutien permanent, leur amour et l'attention qu'ils ont toujours portés à mon travail.

Je remercie Sylvie, Alain, Léna et Cathel pour leur générosité et l'accueil qu'ils m'ont réservé.

Je témoigne ma profonde reconnaissance à Pauline, pour sa présence quotidienne à mes côtés, pour sa patience, pour son soutien de tous les instants, pour l'attention et l'affection qu'elle m'a portées tout au long de ce travail. Je lui adresse également tous mes encouragements pour son propre travail de thèse de Doctorat déjà bien engagé, et je tiens à lui témoigner tout mon amour !

Table des matières

INTRODUCTION GENERALE	10
1. <i>Contexte général</i>	10
2. <i>Contexte de recherche</i>	11
3. <i>Position du problème</i>	11
4. <i>Objectifs et contributions</i>	12
5. <i>Organisation du document</i>	13
CHAPITRE 1. LE REFERENCEMENT WEB.....	14
1. <i>Introduction</i>	14
2. <i>Les moteurs de recherche</i>	15
2.1 Principe de fonctionnement	15
2.2 Classement des résultats.....	16
2.3 Historique des moteurs de recherche	19
3. <i>Principe du référencement web</i>	21
3.1 Comportement des internautes	21
3.2 Définition du référencement web	22
3.3 Types de résultats : naturels et sponsorisés	23
4. <i>Les techniques de référencement web</i>	25
4.1 Compétences requises	25
4.2 Les critères <i>on-page</i>	26
4.3 Les critères <i>off-page</i>	28
4.4 White hat versus black hat	29
5. <i>Les outils du référencement web</i>	31
6. <i>Discussion et conclusion</i>	36
CHAPITRE 2. PROBLEMATIQUE	38
1. <i>Introduction</i>	38
2. <i>Un problème complexe</i>	39
2.1 Structure dynamique et évolutive.....	39
2.2 Volume et diversité des données	41
2.3 Volume et diversité des interactions	42
2.4 Résultats incertains	43
2.5 Bilan	44
3. <i>Le processus de référencement web</i>	45
3.1 Le référencement web : état de la littérature	45
3.2 Absence de cadre	46
3.3 Le référencement web est un processus	46
4. <i>Problématique et hypothèses de travail</i>	47
5. <i>Discussion et conclusion</i>	49
CHAPITRE 3. LE PROCESSUS DE REFERENCEMENT WEB.....	50
1. <i>Introduction</i>	50
2. <i>Méthode pour définir le processus de référencement web</i>	51
2.1 Similitude avec la conception de produits/services.....	51
2.2 Utilisation d'un méta-modèle d'ingénierie	52
2.3 Identification du processus de référencement web	54

Table des matières

2.4	Bilan	56
3.	<i>Acteurs et éléments du processus de référencement web</i>	57
3.1	Les acteurs humains	57
3.1.1	Le référenceur	57
3.1.2	Le client	57
3.2	Les étapes du processus de référencement web	58
3.2.1	Étape 1 : Identification du besoin	58
3.2.2	Étape 2 : Étude de la faisabilité du projet	59
3.2.3	Étape 3 : Proposition des actions	59
3.2.4	Étape 4 : Implémentation des actions	60
3.2.5	Étape 5 : Mesure de l'impact	60
3.3	Interactions entre les acteurs du processus de référencement web	62
4.	<i>Collaboration et niveau de connaissance</i>	63
4.1	Deuxième version du processus de référencement web	65
5.	<i>Discussion et conclusion</i>	68
CHAPITRE 4. MODELISATION DU PROCESSUS DE REFERENCEMENT WEB		70
1.	<i>Introduction</i>	70
2.	<i>Le paradigme agent pour modéliser le processus de référencement web</i>	71
2.1	Le référencement web : un système distribué	71
2.2	Choix de l'outil de modélisation : les systèmes multi-agents	72
3.	<i>Un modèle d'agents du processus de référencement web</i>	74
3.1	Le modèle	74
3.2	Modèle des agents SEO	76
3.3	Propriété des agents SEO	76
3.4	Comportement d'un agent SEO	77
3.5	Coopérations et interactions	78
4.	<i>Modélisation agents du processus de référencement web</i>	80
4.1	Organisation des agents	82
4.1.1	Les agents besoin	82
4.1.2	Les agents fonction	82
4.1.3	Les agents contrainte	82
4.1.4	Les agents solution	82
4.2	Des agents solution : Page et Critère	83
5.	<i>Discussion et conclusion</i>	86
CHAPITRE 5. CONCEPTION D'UN SYSTEME D'AIDE A LA DECISION POUR LE REFERENCEMENT WEB		88
1.	<i>Introduction</i>	88
2.	<i>Vers un système d'aide à la décision pour le référencement web</i>	89
2.1	Objectif de notre système d'aide à la décision	89
2.2	Caractéristiques de notre système d'aide à la décision	91
3.	<i>Un système pour la simulation du processus de référencement web</i>	93
3.1	Calcul du classement	94
3.2	Application	96
3.2.1	Définition du projet	96
3.2.2	Simulation	97
4.	<i>Notre système d'aide à la décision</i>	99
4.1	Architecture de notre système d'aide à la décision	99
4.2	Étape 1 : Identifier l'impact des critères	102
4.2.1	Identification de l'impact des critères indépendants	103

Table des matières

4.2.1.1 Critère KWrText.....	104
4.2.1.2 Critère KWrTitle	105
4.2.1.3 Critère KWrH1	105
4.2.1.4 Bilan des expérimentations conduites sur des critères indépendants	106
4.2.2 Identification de l'impact des critères combinés.....	107
4.2.2.1 Bilan des expérimentations conduites sur des critères combinés	109
4.3 Étape 2 : Définition des ensembles flous.....	109
4.3.1 Bilan sur la définition des ensembles flous.....	111
4.4 Étape 3 : Définition des règles de décision floues	112
4.4.1 Bilan sur la définition des règles floues.....	114
5. <i>Présentation du système d'aide à la décision</i>	114
5.1 Essai et validation du système en conditions réelles	120
6. <i>Discussion et conclusion</i>	123
CONCLUSION GENERALE.....	125
1. <i>Le référencement web est un processus complexe et structurable : apports et limites</i> ..	126
2. <i>Le processus de référencement web est modélisable : apports et limites</i>	127
3. <i>Une meilleure maîtrise du référencement web est possible grâce à un système d'aide à la décision : apports et limites</i>	128
PERSPECTIVES ENVISAGEES	130
1. <i>Vers des systèmes intelligents et coopératifs pour assister l'ingénierie du référencement</i> <i>130</i>	
2. <i>Vers une approche multi-critères versus critères forts</i>	133
3. <i>Vers l'intégration des contraintes dynamiques de coût et de temps passé</i>	134
BIBLIOGRAPHIE.....	135
Liste de nos publications sur le travail conduit	141

Table des illustrations

Figure 1. Étapes du fonctionnement d'un moteur de recherche	16
Figure 2. Archie, le premier moteur de recherche web, 1990	20
Figure 3. Le Triangle D'or, identifié en 2005 par Did-it, Enquiro, and Eyetools durant une étude d'eye-tracking sur le moteur de recherche Google	22
Figure 4. Différence entre les résultats naturels et les résultats sponsorisés sur le moteur de recherche Google avec la requête : « référencement »	24
Figure 5. Interface de l'outil de suivi de positionnement, MyPoseo	32
Figure 6. Interface de l'outil d'audit SEO, affichage des balises <title>, ScreamingFrog	33
Figure 7. Interface de l'outil d'analyse de liens retour, Ahrefs	34
Figure 8. Interface de l'outil d'analyse de mots-clés, SemRush	35
Figure 9. Interface du service Whois pour le nom de domaine utbm.fr, whois.domaintools.com	36
Figure 10. Sites web récompensés par le filtre Google Panda ainsi qu'en témoigne l'évolution positive de leurs visibilitées, 2011, raking-metrics.fr	40
Figure 11. Sites web pénalisés par le filtre Google Panda ainsi qu'en témoigne l'évolution négative de leurs visibilitées, 2011, raking-metrics.fr	40
Figure 12. Trois types d'évolution de classement possibles : (A) augmentation, (B) stagnation et (C) régression	43
Figure 13. Première version du processus de référencement	55
Figure 14. Exemple d'un premier rapport de positionnement, "rapport point zéro", MyPoseo	59
Figure 15. Diagramme des cas d'utilisation, processus de référencement	61
Figure 16. Deuxième version du processus de référencement, cas d'un client expérimenté	66
Figure 17. Diagramme des cas d'utilisation, seconde version du processus de référencement collaboratif	68
Figure 18. Interactions entre acteurs et entités du système	72
Figure 19. Modèle d'agents du processus de référencement	75
Figure 20. Architecture fonctionnelle d'un agent SEO	78
Figure 21. Représentation d'un échange entre deux agents SEO (double interaction)	79
Figure 22. Modélisation agents du processus de référencement web	81
Figure 23. Agents solution : Page, Mot-clé et Backlink	83
Figure 24. Diagramme de séquence représentant l'échange de messages entre agents et acteurs humains durant le processus de référencement	85
Figure 25. Trois types de systèmes d'aide à la décision : SAD guidé par un modèle, SAD guidé par les connaissances et SAD hybride	92
Figure 26. Représentation graphique pour nos trois cas de figure	95
Figure 27. Différence entre Page pharmacie et Page non-pharmacie	97
Figure 28. Simulateur de référencement multi-agents (environnement NetLogo), affichage du classement des six sites de pharmacies	98
Figure 29. Diagramme de classes UML du processus de référencement	99
Figure 30. Diagramme de classes UML représentant l'agent solution Page	100
Figure 31. Diagramme de séquence représentant l'échange de messages entre agents solution et acteurs humains durant le processus de référencement	100

Table des illustrations

Figure 32. Évolution du classement au cours du temps pour le couple de critères n°2 : KWrText et KWrH1, interface Myposeo.....	107
Figure 33. Classement stabilisé des 24 sites web sur le couple de critères n°1 : KWrText et KWrTitle	108
Figure 34. Classement stabilisé des 24 sites web sur le couple de critères n°2 : KWrText et KWrH1	109
Figure 35. Trois ensembles flous pour le critère KWrText : bad, medium et excellent, Matlab	110
Figure 36. Trois ensembles flous pour le critère KWrTitle : bad, medium et excellent, Matlab	111
Figure 37. Trois ensembles flous pour le critère KWrH1 : bad, medium et excellent Matlab.....	111
Figure 38. Règles floues pour les critères KWrText et KWrTitle, éditeur de règle MatLab	112
Figure 39. Graphique de surface du classement pour les critères KWrText et KWrTitle, Matlab	113
Figure 40. Graphique de surface du classement pour les critères KWrText et KWrH1, Matlab	113
Figure 41. Architecture du système d'aide à la décision pour le référencement	115
Figure 42. Définition des ensembles flous pour les critères KWrText, KWrH1 et KWrTitle au format .fcl	116
Figure 43. Définition des règles de décision floues pour les critères KWrText et KWrTitle	117
Figure 44. Interface expert du système d'aide à la décision	117
Figure 45. Interface standard du système d'aide à la décision	118
Figure 46. Deuxième version du système d'aide à la décision, interface expert	119
Figure 47. Deuxième version du système d'aide à la décision, interface standard	119

Liste des tableaux

Tableau 1. Méta-modèle d'ingénierie issu de la conception de produits, [47]	54
Tableau 2. Comparatif sur l'adéquation d'outils courants de modélisation de systèmes aux caractéristiques du processus de référencement	73
Tableau 3. Table descriptive des actes de coopérations des agents SEO	79
Tableau 4. Table des messages	86
Tableau 5. Exemples de valeurs pour le critère c_8 =longueur de la balise <title>	94
Tableau 6. Équations proposées pour nos trois cas de figure	95
Tableau 7. Valeurs obtenues pour les critères de chacun des six sites de pharmacie sélectionnés....	96
Tableau 8. Outils utilisés par le système d'aide à la décision	115
Tableau 9. Premier questionnaire.....	120
Tableau 10. Résultat des tests de diagnostic SEO avec et sans l'outil	121
Tableau 11. Deuxième questionnaire.....	121

Introduction générale

1. Contexte général.....	10
2. Contexte de recherche.....	11
3. Position du problème	11
4. Objectifs et contributions	12
5. Organisation du document	13

1. Contexte général

Le développement rapide du World Wide Web depuis les années 1990 a entraîné l'accroissement du nombre de sites web et l'intensification de la concurrence. Parallèlement, les moteurs de recherche ont joué un rôle majeur en facilitant la recherche d'informations sur le web, au point de devenir une source inépuisable de trafic dont les propriétaires de sites web souhaitent profiter [1], [2]. Afin de capter un maximum de trafic qualifié, il est devenu nécessaire de se démarquer pour améliorer la visibilité d'un site web sur les moteurs de recherche. Pour cela les propriétaires de sites web peuvent faire appel à des entreprises pratiquant le référencement, et plus particulièrement le **référencement naturel** qui regroupe l'ensemble des techniques d'optimisation permettant d'indexer et de positionner les pages d'un site web au plus haut niveau dans les résultats de recherche [3].

L'activité de référencement est conduite par le **référéncieur** qui, en optimisant un certain nombre de critères au sein du site web, améliore son positionnement. Mais les résultats obtenus sont **incertains**, car les algorithmes de classement des moteurs de recherche sont en perpétuelle mutation. Dès lors, il est nécessaire pour le référencieur d'avoir une meilleure maîtrise du processus de référencement et d'être aidé dans ses prises de décisions afin de satisfaire ses clients.

Nos travaux s'inscrivent dans ce cadre et ont ainsi pour objectif de contribuer à la conception et à la maîtrise du processus de référencement web.

2. Contexte de recherche

Nos travaux de recherche se sont déroulés dans le cadre d'une convention CIFRE (Conventions Industrielles de Formation par la REcherche) signée entre le laboratoire IRTES (Institut de Recherche sur les Transports, l'Énergie et la Société) et l'entreprise ORIXA MEDIA¹ que nous appellerons par la suite « entreprise pilote », dont l'activité principale est le référencement de sites web. Ce contexte particulier² nous a permis d'étudier l'environnement et les méthodes de travail des référenceurs en conditions réelles, et nous a permis de mettre en évidence **des lacunes dans la maîtrise du processus de référencement**. Cette proximité avec le monde de l'entreprise nous a également permis d'adapter notre raisonnement et nos approches en élaborant des connaissances nouvelles, voire des outils pour mieux répondre aux besoins des référenceurs et aux attentes de leurs clients.

3. Position du problème

Le processus de référencement est **complexe**, les résultats sont **longs à obtenir** et ils sont **incertains**. Par ailleurs, les moteurs de recherche communiquent très peu sur le **fonctionnement, l'évolution et les mutations de leurs algorithmes** de classement des pages, ce qui complique le travail des référenceurs. Qui plus est, ils doivent respecter des **impératifs de temps et de coûts**. Ils travaillent aussi beaucoup **par expérience** et de manière plus ou moins **artisanale** du fait de l'absence de cadre précis pour l'exercice de cette activité [4]. Ils doivent également concilier les besoins de leurs clients et le respect des règles des moteurs de recherche qui sont parfois en contradiction. C'est la raison pour laquelle il est important de **collaborer avec le client** [5] en amont du projet de référencement. Dès lors, il est nécessaire de définir un réel cadre à cette activité en proposant un **proces-**

¹ SARL *Orixa Media* au CA de 2 732 632€ (en 2011) spécialisée dans le référencement de sites web. Siège basé à Paris (11^e) et filiale *Visiperf* à Montbéliard (25200).

² Les semaines de travail étaient divisées entre une période en entreprise (70%) et une période en laboratoire de recherche (30%)

sus de référencement prenant en compte les attentes et les besoins des clients tout en respectant les règles des moteurs de recherche. Par ailleurs, afin de permettre une **meilleure maîtrise de l'incertitude** et ainsi faciliter le travail du référenceur, celui-ci doit être guidé dans ses choix et actions à mener. Le développement d'un **système d'aide à la décision** adapté à l'activité du référenceur et servant de support pour le bon déroulement d'un projet de référencement est une des réponses possibles à notre problème comme nous avons pu le montrer [6]. C'est en particulier sur cet axe que nous avons choisi de porter nos travaux de recherche.

4. Objectifs et contributions

Le premier objectif de cette thèse est de permettre une **meilleure compréhension du processus de référencement**. Pour cela, nous nous proposons de **définir, modéliser et concevoir le processus de référencement** dans son ensemble, de la prise de contact avec le client jusqu'à la finalisation du projet de référencement.

Le deuxième objectif est de permettre la **maîtrise du processus de référencement** par le couple référenceur/client. Grâce à la modélisation du processus et à partir des connaissances acquises précédemment, nous nous sommes proposé de **développer un système d'aide à la décision** permettant à la fois une meilleure maîtrise du référencement et jouant un véritable rôle de support dans le choix des actions à mener.

Ainsi, pour résumer :

- La première contribution de ce travail consiste à établir un cadre efficace pour l'activité de référencement en définissant les différentes **étapes du processus de référencement** grâce à la littérature, l'expérience en entreprise et l'utilisation d'un méta-modèle d'ingénierie.
- Après avoir défini ces différentes étapes, la deuxième contribution consiste à modéliser par un **système multi-agents** le processus de référencement afin de permettre une meilleure compréhension de son fonctionnement et de contrôler efficacement ses interactions.

- La troisième contribution porte sur la **conception d'un système d'aide à la décision** développé à partir de notre modèle d'agents et utilisant la logique floue. Il est conçu pour servir de support à l'activité de référencement. Afin de valider de notre modèle, nous avons testé ce système d'aide à la décision en condition réelle (dans l'entreprise pilote).

5. Organisation du document

Notre document est organisé en six chapitres :

Chapitre 1. Le référencement web

Chapitre 2. Problématique

Chapitre 3. Le processus de référencement web

Chapitre 4. Modélisation du processus de référencement web

Chapitre 5. Conception d'un système d'aide à la décision pour le référencement web

Chapitre 6. Conclusion générale et perspectives

Le premier chapitre présente un **état de l'art** sur la notion de référencement web. Le deuxième chapitre expose les problèmes liés à la pratique du référencement web et présente notre **problématique de recherche** ainsi que nos hypothèses de travail. Dans le troisième chapitre, nous proposons de **définir** le processus de référencement ainsi que ses étapes à partir d'un méta-modèle d'ingénierie. Dans le quatrième chapitre, nous **modélisons** le processus de référencement grâce à un système multi-agents. Le cinquième chapitre présente notre **système d'aide à la décision** pour le référencement web conçu à partir de notre modèle d'agents et validé à partir d'une série d'expérimentation et de tests en condition réelle dans l'entreprise pilote. Nous terminons enfin ce travail de recherche par une **conclusion générale** suivie de **perspectives** possibles à partir de nos résultats.

Chapitre 1. Le référencement web

1. Introduction.....	14
2. Les moteurs de recherche.....	15
2.1 Principe de fonctionnement.....	15
2.2 Classement des résultats.....	16
2.3 Historique des moteurs de recherche.....	19
3. Principe du référencement web.....	21
3.1 Comportement des internautes.....	21
3.2 Définition du référencement web.....	22
3.3 Types de résultats : naturels et sponsorisés.....	23
4. Les techniques de référencement web.....	25
4.1 Compétences requises.....	25
4.2 Les critères <i>on-page</i>	26
4.3 Les critères <i>off-page</i>	28
4.4 White hat versus black hat.....	29
5. Les outils du référencement web.....	31
6. Discussion et conclusion.....	36

1. Introduction

Depuis que les moteurs de recherche sont apparus et qu'ils se sont démocratisés, les propriétaires de sites web ont voulu bénéficier de visibilité à travers ce nouveau vecteur de trafic. La pratique du référencement est devenue inévitable pour espérer briguer les meilleures positions sur les moteurs de recherche, ou en tout cas pour s'en rapprocher. Les référenceurs sont les experts qui permettent d'atteindre, dans la majorité des cas, les objectifs de visibilité que se sont fixés les propriétaires de sites web.

Ce chapitre illustre le contexte de notre étude à travers un état de l'art du référencement. Nous y présentons le principe, les techniques d'optimisation et quelques outils couramment utilisés par les référenceurs dans le cadre de leur activité.

L'objectif de ce chapitre est de **capitaliser les connaissances sur l'activité de référencement**. Ceci nous permettra d'aborder convenablement les prochains chapitres dédiés à la problématique, à la conception puis à la maîtrise du processus de référencement.

La première section introduit ce chapitre par une synthèse sur les moteurs de recherche et explique leur principe de fonctionnement. La deuxième section définit le principe de référencement web. La troisième section présente les techniques de référencement web ainsi que les différents critères à améliorer. La quatrième section expose quelques outils couramment utilisés par les référenceurs pour exercer leur activité. Enfin nous terminerons ce chapitre par une discussion et une conclusion.

2. Les moteurs de recherche

Avant d'énoncer le principe de référencement web, nous allons tout d'abord présenter le concept de moteur de recherche, car ces deux notions sont intimement liées. Pour maîtriser l'activité de référencement, le référenceur doit avant tout comprendre le fonctionnement des moteurs de recherche et le comportement de leurs algorithmes de classement. Cette section sert de préambule à nos travaux de recherche, nous y présentons le fonctionnement, les différentes méthodes de tri et l'histoire des moteurs de recherche.

2.1 Principe de fonctionnement

Les moteurs de recherche tels que nous les connaissons aujourd'hui, bien que distincts sur leurs méthodes de classement, fonctionnent tous de manière assez similaire. Leur fonctionnement peut être décomposé en quatre étapes [7]:

Étape 1 : La collecte d'information

Cette étape permet de rassembler et de collecter (*crawl*) les informations provenant des différentes pages web. Cette étape est réalisée par des robots d'indexation (*crawlers* ou *spiders*).

Étape 2 : L'indexation des pages

Cette étape permet d'enregistrer les données collectées durant l'étape précédente au sein de bases de données (*index*).

Étape 3 : Le traitement des requêtes

Pendant cette étape, les données sont traitées en fonction des requêtes saisies par l'internaute. Les méthodes et les critères utilisés par le moteur de recherche déterminent le classement final des résultats.

Étape 4 : La restitution des résultats

Cette dernière étape consiste à afficher les résultats du classement effectué lors de l'étape précédente. Les résultats sont couramment affichés sous forme de pages (*SERP : Search Engine Results Pages*) constituées de plusieurs liens redirigeant vers des pages web.

La figure 1 ci-dessous illustre les différentes étapes de fonctionnement d'un moteur de recherche :

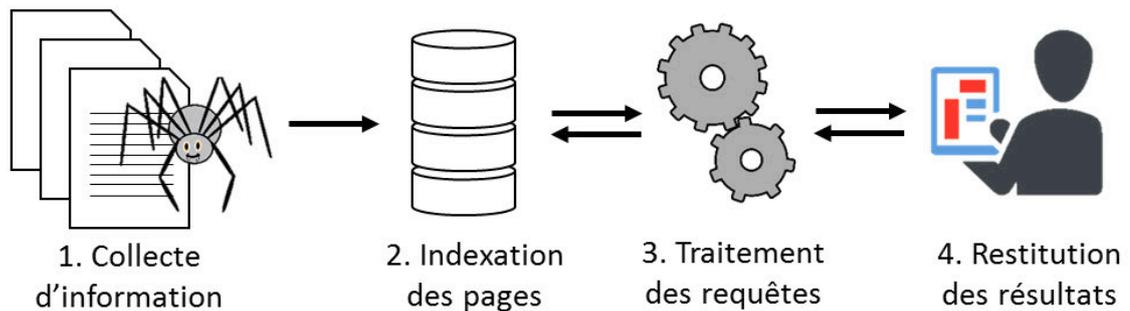


Figure 1. Étapes du fonctionnement d'un moteur de recherche

Le fonctionnement d'un moteur de recherche est donc basé sur le principe suivant : les informations sont d'abord collectées, puis indexées ; elles sont ensuite traitées et enfin restituées à l'utilisateur. Si la majorité des moteurs de recherche fonctionnent en suivant ce principe [1], [2], ils restent tout de même distincts sur la partie traitement et classement des données. En effet, c'est cette partie qui permet à un moteur de recherche de contrôler la pertinence de ses résultats.

2.2 Classement des résultats

La qualité principale d'un moteur de recherche réside dans l'efficacité de son système de classement des résultats (*ranking*) [8]. C'est en fait un véritable facteur différenciant, souvent considéré comme un secret industriel ou protégé par un brevet. Le moteur de recherche doit pouvoir répondre de manière efficace et pertinente à la requête de l'internaute, ceci en un minimum de temps et en consom-

mant le moins de ressources possible. C'est d'ailleurs la raison pour laquelle la mutation des algorithmes de classement est permanente³. Sans cela, le déclin du moteur de recherche serait inéluctable, ce qui s'est déjà produit⁴. Bien que les différents algorithmes de classement soient rarement rendus publics, certaines méthodes de classement ont démontré leur efficacité. D'ailleurs, les expérimentations que nous avons pu conduire comme nous le verrons ont prouvé que certaines d'entre elles sont encore utilisées par les moteurs de recherche aujourd'hui. Nous présentons ci-dessous les quatre principales méthodes de tri utilisées par les moteurs de recherche aujourd'hui.

➤ Le tri par pertinence

La méthode de tri par pertinence est utilisée par la plupart des moteurs de recherche [9], comme nous le verrons dans la section 4. Les référenceurs en ont d'ailleurs bien conscience ce qui nous a amené à réutiliser certains de ces critères dans nos expérimentations. Le tri par pertinence permet d'évaluer le classement d'une page selon cinq facteurs :

- **Le poids d'un mot dans un document** : la localisation du mot dans la page influence son classement. Par exemple, le moteur de recherche peut considérer que la présence du mot-clé dans l'en-tête de la page aura plus de poids que dans le pied de page.
- **La densité d'un mot** : la fréquence de répétition du mot-clé par rapport au nombre total de mots dans la page influence son classement. Par exemple, un mot-clé répété 8 fois dans un texte de 250 mots équivaut à une densité de 3,2%. Plus la densité est élevée, meilleur sera le classement de la page.

³ Le moteur de recherche Google effectue 500 changements d'algorithme par an afin de toujours proposer des résultats pertinents, <https://youtu.be/J5RZOU6vK4Q>

⁴ Suite à la perte de qualité de ses résultats, le 8 juillet 2013, Yahoo annonce la fermeture du moteur de recherche AltaVista, <https://yahoo.tumblr.com/post/54125001066/keeping-our-focus-on-whats-next>

- **Le poids d'un mot dans la base** : la fréquence de répétition d'un mot-clé dans l'ensemble de la base de données influence son classement. Par exemple, les mots-clés « le » ou « la » seront défavorisés ou éliminés, car ils sont très fréquents.
- **La correspondance d'expression** : la similarité entre l'expression recherchée par l'internaute et l'expression correspondante dans la page influence son classement. Sur les requêtes à plusieurs mots-clés, cela est un gage de pertinence.
- **La relation de proximité** : la proximité entre les termes de la requête et ceux contenus dans la page influence son classement. Plus les termes sont proches, plus la page sera favorisée.

➤ **Le tri par popularité**

Le tri par popularité favorise les pages web possédant un grand nombre de liens retour (*backlinks*). Cette méthode a permis de contrer les abus constatés sur les méthodes de tri par pertinence. Plus une page possède de liens provenant d'autres pages, plus son classement dans les résultats sera favorisé. Cette méthode de tri a été utilisée par Google en 1998 pour créer l'algorithme PageRank [10].

➤ **Le tri par mesure d'audience**

Cette méthode de tri a été créée par la société Direct Hit en 1998. Le tri par mesure d'audience favorise les pages web, dont la « quantité » et la « qualité » des visites sont la plus élevées. Une page bien classée justifie donc d'un nombre de visites important, mais elle doit aussi satisfaire l'internaute. En examinant par exemple la durée des visites sur une page, il est possible de juger de sa qualité, et donc d'influencer son classement.

➤ **Le tri par catégorie**

Créé par la société Northernlight en 1997, la méthode de tri par catégorie (*clustering*) classe les pages web en fonction de leur thématique. Cette technique utilise donc la sémantique d'une page pour déterminer son classement.

➤ Autres méthodes de tri

D'autres méthodes de tri ont été développées depuis afin de proposer des résultats toujours plus pertinents. Nous pouvons citer l'algorithme HITS (*Hyperlink-Induced Topic Search*) [11] qui privilégie les sites web faisant des liens vers des sites d'influence (appelés *authorities*). On peut aussi citer le Trustrank [12] qui a pour objectif de favoriser les sites web de « confiance ». Cette méthode de tri s'est perfectionnée avec l'ajout de nombreux critères de confiance (ex : ancienneté du nom de domaine, niveau de sécurisation du site, etc.). Cela a donné lieu à plusieurs dépôts de brevet [13], [14]. Ask.com de son côté a développé l'algorithme ExpertRank⁵ qui considère la « popularité thématique » : il privilégie les liens provenant de sites à thématique similaire. Enfin l'algorithme BrowseRank [15] développé par Microsoft étudie le comportement des internautes lors de leur visite (ex : durée de la visite, taux de rebond, etc.)

En conclusion, les différentes méthodes de tri présentées, bien que relativement anciennes, sont encore aujourd'hui utilisées par la plupart des moteurs de recherche. Si le moteur de recherche Google utilisait uniquement le tri par popularité en 1998, aujourd'hui il a beaucoup évolué et utilise désormais une large gamme de méthodes de tri, dont celle que nous avons présentée.

2.3 Historique des moteurs de recherche

Après avoir étudié le fonctionnement d'un moteur de recherche et leurs méthodes de classement, nous présentons dans cette section un historique de leur apparition. Le premier moteur de recherche web dénommé *Archie*⁶ a été créé en 1990 par Alan Emtage et J. Peter Deutsch de l'Université McGill à Montréal. Il utilisait le protocole FTP et permettait d'accéder à un listing de fichiers distants. Le Centre Interdisciplinaire de Modélisation Mathématique et Informa-

⁵ <http://sp.fr.ask.com/fr/docs/about/asksearch.shtml>

⁶ <http://people.ischool.illinois.edu/~chip/projects/timeline/1990archie.htm>

tique de l'Université de Varsovie maintient encore aujourd'hui une version du serveur Archie à des fins historiques comme l'illustre la figure 2 qui suit :



Figure 2. Archie, le premier moteur de recherche web, 1990

Les moteurs HTTP que nous connaissons à l'heure actuelle ont vu le jour en 1993 avec l'arrivée de *Wanderer*⁷ qui proposait le tout premier système de robot d'indexation. *Aliweb* [16], présenté un an plus tard regroupait les technologies d'Archie et de *Wanderer* au sein d'une même application. Il permettait de rechercher du contenu parmi des index de données fournis manuellement par les webmasters. Toujours en 1994, Yahoo!⁸ est lancé, mais ce n'est encore qu'un annuaire web, il ne deviendra moteur de recherche que huit ans plus tard, en 2002. En 1995, c'est l'arrivée d'*AltaVista* qui est le premier à utiliser un *crawler* permettant l'indexation rapide des pages web. Il fut également le premier moteur de recherche multilingue. En 1998, *Google* est lancé⁹, la pertinence et la rapidité de ses résultats lui ont permis de gagner la confiance de ses utilisateurs et d'acquérir rapidement

⁷ <http://www.mit.edu/people/mkgray/growth/>

⁸ http://www.leepublicschools.net/Technology/Search-Engines_Directories.pdf

⁹ Google est aujourd'hui le moteur de recherche le plus utilisé en France. En juin 2016, les parts de marchés étaient les suivantes: Google 92,55%, Bing à 4,45% et Yahoo! à 2,06%, <http://gs.statcounter.com>

une place de leader sur le marché très concurrentiel des moteurs de recherche.

Cette section nous a permis de mieux comprendre le fonctionnement, le comportement et l'historique des moteurs de recherche. Nous retiendrons que pour continuer à exister, un moteur de recherche doit sans cesse évoluer et se perfectionner afin de fournir des résultats toujours plus pertinents.

3. Principe du référencement web

Dans la section précédente, nous avons étudié le concept de moteur de recherche. Dans cette troisième section, nous abordons le principe de référencement. Une première étude portera sur le comportement des internautes, puis nous définirons le concept de référencement. Enfin nous présenterons les deux principaux types de référencement existant pour ensuite conclure cette section.

3.1 Comportement des internautes

Pour assimiler correctement le principe du référencement, il est nécessaire, voire indispensable, de connaître le comportement des internautes. En juin 2013, une étude a montré que **54% des internautes utilisaient un moteur de recherche** pour effectuer une recherche sur le web (contre 50% en 2011)¹⁰. La majorité du trafic sur le web provient ainsi des moteurs de recherche, ce qui traduit l'intérêt pour les propriétaires de sites web d'être présents sur ce type de support. Par ailleurs, en 2011 une autre étude montre que le **taux de clic sur la première page était de 53%** (contre 49% en 2006)¹¹. Dès lors, avoir son site présent sur un moteur de recherche ne suffit plus, il est aussi nécessaire d'être visible sur la première page. Enfin en 2012, il est montré que 56% des clics sont récoltés par le premier résultat, 15% par le deuxième, et 9% par le troisième¹². Ainsi, **80% de clics géné-**

¹⁰ <https://www.forrester.com/report/How+Consumers+Found+Websites+In+2012/-/E-RES92661>

¹¹ <http://www.abondance.com/actualites/20120103-11216-etude-sur-le-positionnement-et-le-taux-de-clics-dans-les-resultats-des-moteurs-de-recherche.html>

¹² <http://success.compete.com/seeing-between-the-lines-of-the-search-and-the-click-whitepaper>

rés se concentrent sur les trois premières positions. La visibilité sur la première page ne suffit plus, pour espérer acquérir un maximum de trafic un site web doit être visible dans les trois premières positions. C'est ce qu'on appelle le « Triangle d'Or » représenté par la figure 3 ci-dessous, où la visibilité d'un site atteint les 100% [17]. Selon cette étude, la zone rouge vif en haut à gauche sur la figure représente la zone la plus lue instinctivement par l'œil humain.

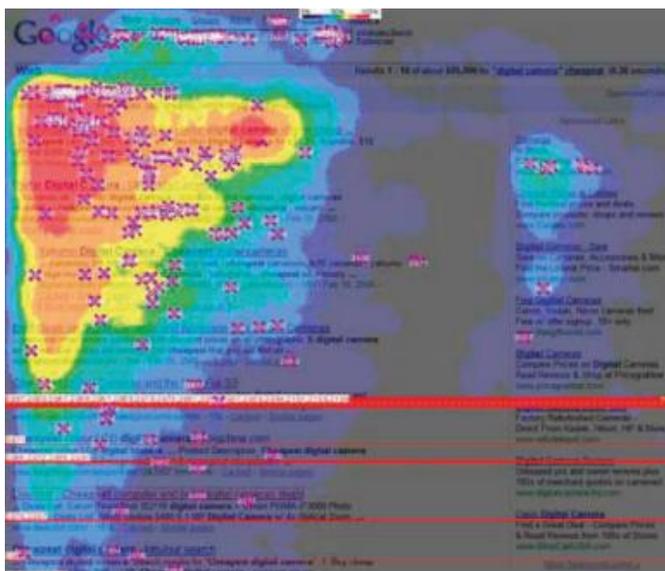


Figure 3. Le Triangle D'or, identifié en 2005 par Did-it, Enquiro, and Eyetools durant une étude d'eye-tracking sur le moteur de recherche Google

Ces différentes études permettent de mieux comprendre pourquoi les propriétaires de sites web souhaitant acquérir davantage de trafic doivent impérativement améliorer la position et donc la visibilité de leur site sur les moteurs de recherche. La pratique de l'activité de « référencement » est une des réponses à ce besoin de visibilité.

3.2 Définition du référencement web

Avant de définir le référencement web, nous proposons de lever une confusion courante concernant les deux termes suivants : « *référencement* » et « *positionnement* ».

À l'origine le « *référencement* » comprend « les actions nécessaires à l'indexation et à l'enregistrement de pages web dans la base de données d'un moteur de recherche ». Le « *positionnement* » quant à lui regroupe « les techniques d'optimisation nécessaires à l'amélioration

du classement des pages dans les résultats de recherche ». Mais l'analogie entre ces deux termes est devenue très courante, c'est notamment le cas au sein de l'entreprise dans laquelle nous avons effectué nos travaux de recherche. Par souci de simplification, nous utiliserons le terme référencement pour définir « **le travail d'indexation et d'optimisation des pages web dans l'optique d'améliorer leur position et leur visibilité dans les résultats de recherche** ».

En anglais, le terme « *positionnement* » se traduit « *ranking* ». Le terme « *référencement* » quant à lui peut se traduire par « *search engine marketing* » mais les anglo-saxons préfèrent utiliser les termes « *Search Engine Optimization* » (référencement naturel) et « *Search Engine Advertising* » (référencement sponsorisé) [18] qui permettent de faire la distinction entre deux types de référencement différents dont nous expliquerons le principe dans la section suivante.

3.3 Types de résultats : naturels et sponsorisés

Un site web peut être référencé de deux manières. Nous définirons dans un premier temps ce qu'est le référencement dit « *naturel* » ou « *Search Engine Optimization* » souvent abrégé « *SEO* » en anglais. Afin de bien comprendre sa signification, nous devons distinguer les résultats « *naturels* » et les résultats « *sponsorisés* » avec qui l'amalgame est fréquent. La figure 4 qui suit, qu'il est facile d'obtenir en tapant le terme « référencement » (sans accents), présente la différence visuelle entre les résultats « *naturels* » et les résultats « *sponsorisés* » sur le moteur de recherche Google.fr.

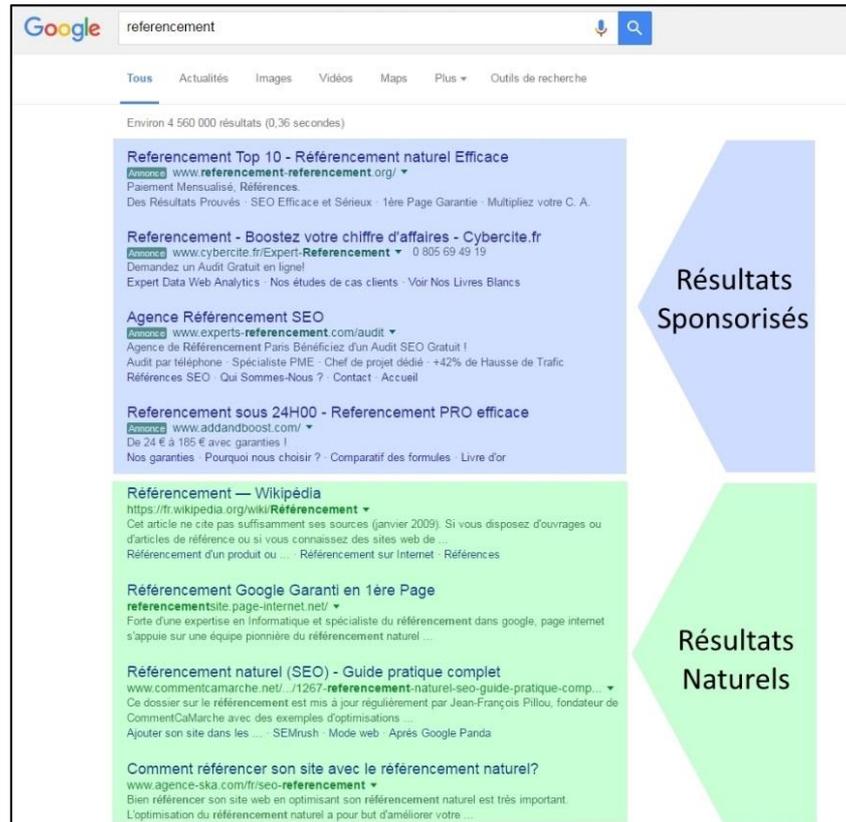


Figure 4. Différence entre les résultats naturels et les résultats sponsorisés sur le moteur de recherche Google avec la requête : « référencement »

Les résultats « *sponsorisés* » correspondent aux quatre liens affichés dans la zone bleue qui est en fait une zone publicitaire. Si un annonceur souhaite être visible à cet emplacement, il devra payer au moteur de recherche chaque clic effectué sur ses annonces. Le coût du clic est déterminé par un système d'enchère, ainsi plus le mot-clé sera concurrentiel, plus le prix du clic sera élevé [18].

Les résultats « *naturels* » correspondent aux résultats visibles dans la zone verte. Ils sont déterminés par le classement issu de l'algorithme du moteur de recherche selon la ou les méthodes de traitement choisies par celui-ci (ex : tri par pertinence, tri par popularité, etc.) [3].

Bien qu'affichés sur la même page, les résultats « *naturels* » et « *sponsorisés* » ne fonctionnent pas selon le même principe. Par ailleurs, les deux types de résultats sont indépendants. Rappelons que la survie d'un moteur de recherche est grandement liée à la pertinence et à l'objectivité de ses résultats, il est donc nécessaire pour un moteur de recherche de fournir des résultats indépendants des budgets publicitaires [7].

Nous retiendrons de cette section que pour augmenter le trafic d'un site web, il est nécessaire d'apparaître sur la première page des moteurs de recherche, idéalement dans les trois premiers résultats qui sont les plus consultés. Le référencement, et notamment le référencement « *naturel* » est un moyen d'améliorer la position d'un site web pour y attirer davantage de trafic.

4. Les techniques de référencement web

Dans la section précédente, nous avons montré l'intérêt pour un propriétaire de site web de faire du référencement. Dans cette section nous étudierons comment pratiquer l'activité de référencement. Nous définirons la notion de « *critère* » en donnant quelques exemples parmi différentes catégories : « *on-page* » et « *off-page* ». Enfin nous expliquerons la différence entre « *white hat* » et « *black hat* » afin de bien discerner ces deux types de méthodes pour pratiquer le référencement naturel.

4.1 Compétences requises

Pour référencer un site web, il est nécessaire d'avoir des connaissances en programmation web et de maîtriser les plateformes de gestion de contenu (Content Management System), au moins les plus fréquentes¹³ telles que *Wordpress*, *Joomla!*, *Drupal*, *Prestashop* et *Magento*. Ensuite il est nécessaire d'avoir une bonne connaissance des critères de classement, ou au moins des plus impactants.

Le métier de référenceur est très récent puisqu'il s'est créé avec l'arrivée des moteurs de recherche web au milieu des années 1990. Si les premiers référenceurs étaient pour la grande majorité des autodidactes, désormais des programmes de formations existent¹⁴ ce qui permet la légitimation de la profession. Domenget considère ce métier comme pérenne, mais toujours en quête de légitimité [4]. Il a pu

¹³ <https://trends.builtwith.com/cms>

¹⁴ La première formation de Référenceur & Rédacteur Web (RRW) a été créée en 2008 à l'Université de Haute Alsace à Mulhouse

mettre en évidence cinq profils types de référenceurs : les Défricheurs, les Explorateurs, les Organisateurs, les Planificateurs et les Stratèges, distinct de par leur parcours, mais aussi de par leur vision de cette activité.

4.2 Les critères *on-page*

Même si la connaissance des algorithmes de classement est limitée pour des raisons de confidentialité évidentes, certains critères sont tout de même connus pour influencer positivement la position d'une page web [19]. Nous avons sélectionné les critères les plus impactant en les regroupant dans deux catégories, les critères *on-page* et les critères *off-page*.

Les critères *on-page* sont littéralement « *dans la page* », c'est-à-dire qu'ils peuvent toucher le code source, le contenu ou la structure de la page. Plusieurs chercheurs [20]–[25] ont démontré l'utilité d'optimiser les critères *on-page*, notamment afin d'améliorer le classement d'une page sur un mot-clé spécifique. Voici une liste de critères à fort impact issus de la littérature et couramment utilisés par les référenceurs dans l'entreprise pilote :

- **La balise <title>** : la balise <title> est encore à l'heure actuelle l'un des critères *on-page* les plus impactant pour le référencement. Pour rappel, le texte de la balise <title> s'affiche dans l'en-tête du navigateur web lors de la visite d'une page web. Celle-ci doit être complétée et elle doit être unique sur chaque page. Les moteurs de recherche fixent une limite de taille¹⁵ pour cette balise sinon la balise est tronquée dans les pages de résultats. Pour rédiger cette balise, il est important d'utiliser des mots-clés représentatifs du contenu de la page. À noter que le mot-clé le plus à gauche doit être le plus signifi-

¹⁵ Google autorise un maximum de 70 caractères espace compris, <http://www.webrankinfo.com/dossiers/conseils/taille-maximale-balise-title#gref>. La balise <title> est tellement importante que le moteur de recherche Google a fait le choix de générer automatiquement cette balise à partir du contenu de la page si celle-ci est absente ou incorrecte.

catif. Par ailleurs, répéter un mot-clé plus de trois fois dans la balise <title> peut entraîner des pénalités [22]. Voici l'exemple d'une balise <title> correcte pour une page web parlant de « nettoyage des sols ».

```
<title>Nettoyage des sols – Nos services – ProClean</title>
```

- **La balise <h1>** : la balise h1 est tout aussi importante, la différence avec la balise <title> est que la balise <h1> s'affiche directement dans le contenu de la page. Il doit y avoir une seule balise <h1> par page, elle doit être unique et doit correspondre au contenu et à la thématique de la page. À noter que la présence d'une balise <title> et d'une balise <h1> sur la même page a un impact positif sur le classement [22], [23]. Voici un exemple de balise <h1>, toujours pour une page parlant de « nettoyage des sols » :

```
<h1>Nettoyage des sols</h1>
```

- **Le contenu** : le contenu, et notamment le contenu textuel est primordial pour un bon référencement. Nous considérons le contenu texte comme étant le texte visible par l'internaute, présent entre les balises <body> mais hors des balises <hn>. Chaque page web doit contenir au minimum 200 mots. La règle « une thématique par page » doit être respectée pour ne pas perturber les moteurs de recherche. Les mots-clés importants doivent être répétés, une densité comprise entre 2% et 8% est conseillée [26].

Suite à de nombreux abus, certaines balises HTML comme la balise <meta name="keyword"> ne sont plus utilisées par les moteurs de recherche [7]. De plus, il est rare que des référenceurs complètent cette balise, car elle peut renseigner les concurrents sur la liste des mots-clés visés pas une stratégie de référencement. En conclusion, la présence des termes à référencer à la fois dans le contenu texte et dans les différentes balises est primordiale pour espérer apparaître dans les résultats des moteurs de recherche. Nous avons nous-mêmes utilisé ces critères *on-page* pour réaliser des expérimentations que nous présenterons au chapitre 5.

4.3 Les critères off-page

Dans la section précédente, nous avons présenté les critères *on-page* et nous avons démontré l'intérêt de les optimiser pour espérer améliorer le référencement d'un site web. Auparavant ces critères étaient les seuls utilisés pour classer les pages, notamment sur le moteur de recherche AltaVista. Mais en 1998, Google a changé la donne en lançant le concept de tri par popularité. Ainsi, les critères « en dehors de la page » touchant à son environnement se sont démocratisés, car les résultats obtenus étaient jugés plus pertinents. Comme pour les critères *on-page*, plusieurs chercheurs ont constaté un gain de positions en optimisant ce type de critères. Nous en présentons quelques-uns ci-dessous, issus de la littérature et utilisés dans l'entreprise dans laquelle nos travaux de recherche sont conduits :

- **Les backlinks** : le principe est simple, plus un site web acquiert de « liens retour » provenant d'autres sites web, plus sa popularité est jugée importante, son classement sera alors privilégié par le moteur de recherche. Plusieurs chercheurs [19], [21], [27] ont réussi à démontrer qu'en augmentant artificiellement le nombre de liens à destination d'une page web son classement s'améliorait. Cette stratégie se nomme le *netlinking*.
- **L'ancienneté** : la notion de confiance est importante pour les moteurs de recherche. Ainsi un site web présent depuis de nombreuses années sur le même nom de domaine sera considéré comme un site de confiance. Evans [21] a démontré le fort impact de l'ancienneté sur le classement. Par ailleurs, dans l'entreprise dans laquelle nous avons effectué nos recherches, nous avons remarqué une amélioration rapide du classement sur les sites présents depuis de nombreuses années.

Dans cette section nous avons présenté quelques-uns des critères *off-page* utilisés par les moteurs de recherche et démontré l'importance qu'ils avaient pour le référencement d'un site web. Améliorer les critères *on-page* ne suffit pas, l'environnement d'un site web, sa popularité et son ancienneté ont également un impact très fort sur son référencement.

4.4 White hat versus black hat

Dans la section précédente, nous avons présenté les critères permettant d'optimiser le référencement d'un site web. Dans cette section, nous avons répertorié trois types de pratiques du référencement :

- Le premier type est appelé « *white hat SEO* »¹⁶, elle consiste à respecter les consignes des moteurs de recherche en évitant absolument toute manœuvre frauduleuse (ex : « *concevez vos pages en pensant d'abord aux internautes et non aux moteurs de recherche, évitez les astuces destinées à améliorer le classement sur les moteurs de recherche* »). Ces bonnes pratiques sont toutes répertoriées dans un document appelé « Guidelines »¹⁷.
- Le deuxième type est appelé « *black hat SEO* » [28]–[30]. Ce sont des pratiques qui ne sont pas tolérées par les moteurs de recherche. Les sites web qui les emploient s'exposent à des pénalités de classement, voire un bannissement de l'index. Du côté des critères *on-page*, la plupart des abus touchent la manipulation des textes. Par exemple, les contenus copiés ou dupliqués depuis d'autres sites web sont mal perçus par les moteurs de recherche. Ceux-ci sont désormais capables de détecter l'antériorité d'un contenu pour privilégier le classement du précurseur par rapport aux successeurs. D'autres pratiques consistent à répéter un grand nombre de fois un mot-clé à la suite dans un texte (*keyword stuffing*) [31], ou à cacher des mots-clés [32] en utilisant par exemple une police d'écriture blanche sur un fond blanc. Là encore, la plupart des moteurs de recherche détectent ce genre d'abus. Cela s'est notamment produit pour le site web Bmw.de qui utilisait des techniques du *cloaking* [33], [34] pour améliorer son classement. Le site affichait des pages au design très attrayant avec beaucoup d'images aux internautes « humains », et de longs paragraphes de texte très optimisés pour les moteurs de recherche. En fonction,

¹⁶ Pour le référencement, le terme « white hat » peut se traduire par « chapeau blanc », techniques de référencement éthique, qui respectent les règles des moteurs de recherche, par opposition avec « black hat » (chapeau noir), ensemble des techniques prohibées par les moteurs de recherche

¹⁷ Guidelines de Google : <https://support.google.com/webmasters/answer/35769?hl=fr>

de qui visitait le site (humain ou moteur de recherche), le site adaptait l’affichage de son contenu. Pour sanctionner cette pratique, le moteur de recherche Google a pénalisé le site Bmw.de en 2006¹⁸. Des abus sont également possibles pour les critères *off-page*. La création d’un grand nombre de liens retour (backlinks) permettait de duper les moteurs de recherche et d’améliorer son classement [27]. Mais les algorithmes de classement évoluent, face aux abus, les moteurs de recherche ont implanté dans leur algorithme des critères de « qualité » de lien retour. Ce n’est plus uniquement le volume qui est jugé, mais aussi la qualité. Ainsi, un lien retour de bonne qualité doit être issu d’un site ayant une thématique approchant celle du site destinataire. La fréquence de création des liens est également jugée afin d’éviter la création trop rapide de liens artificiels.

- Enfin le dernier type de pratique s’appelle le « *grey hat SEO* », littéralement « chapeau gris » [35]. Il se positionne entre le « *white hat SEO* » et le « *black hat SEO* ». Cette pratique consiste à manipuler une page web tout en respectant au maximum les règles des moteurs de recherche. Les techniques que nous avons exposées dans les parties 4.2 et 4.3 sont considérées comme du « *grey hat SEO* ». Pour optimiser le classement d’un site web, les référenceurs pratiquent le « *grey hat SEO* ».

Nous avons vu que le référencement demandait certaines optimisations, de différents types et à différents niveaux, aussi bien dans la page web qu’en dehors. Ces optimisations sont nécessaires pour la réussite d’un projet de référencement. Néanmoins comme les algorithmes des moteurs de recherche sont en perpétuelle mutation, la modification de ces critères et de leur impact dans le temps n’est pas à proscrire. Le référenceur doit en être conscient pour adapter ses méthodes de travail, c’est ce qui fait l’une des difficultés de ce métier.

¹⁸ <https://www.mattcutts.com/blog/ramping-up-on-international-webspam/>

5. Les outils du référencement web

Dans la section précédente, nous avons étudié les différentes techniques et critères permettant de référencer un site web. Dans cette section nous allons présenter différents types d'outils et services utilisés par les référenceurs durant l'exercice de leur activité. Les outils et services que nous présentons sont d'ailleurs tous utilisés par les référenceurs de l'entreprise pilote.

Pour évaluer la réussite d'un projet de référencement, il est nécessaire pour le référenceur d'utiliser un outil de suivi de positionnement. Ce type d'outils permet de faire un état des lieux (avant/après) et de constater l'évolution du classement d'un site web.

- **Suivi de positionnement** : ce type d'outil permet d'obtenir la position d'un site web pour un mot-clé donné. Dans l'entreprise pilote, l'outil utilisé est *MyPoseo*¹⁹, comme illustré sur la figure 5 ci-dessous. La colonne de gauche répertorie les mots-clés à optimiser, la colonne du milieu indique le moteur de recherche concerné (Google.fr dans notre cas), et la dernière colonne indique les positions du site web pour une date donnée. L'historique des positions est conservé, ainsi il est possible d'afficher la position d'un mot-clé donné à une date antérieure pour constater l'évolution du classement.

¹⁹ D'autres outils de suivi de positionnement existent comme Yooda, Ranks ou encore Raventools

		Moteur	Position
<input type="checkbox"/>	Mots-clés		23 juil. 2016
<input type="checkbox"/>	saut en parachute vitry le françois	 .fr (fr)	4
<input type="checkbox"/>	saut en parachute valenciennes	 .fr (fr)	4
<input type="checkbox"/>	saut en parachute rethel	 .fr (fr)	6
<input type="checkbox"/>	saut en parachute reims prunay	 .fr (fr)	1
<input type="checkbox"/>	saut en parachute pithiviers	 .fr (fr)	3
<input type="checkbox"/>	saut en parachute merville	 .fr (fr)	25
<input type="checkbox"/>	saut en parachute le treport	 .fr (fr)	3
<input type="checkbox"/>	saut en parachute laon	 .fr (fr)	1
<input type="checkbox"/>	saut en parachute ile de france	 .fr (fr)	51
<input type="checkbox"/>	saut en parachute epernay	 .fr (fr)	5
<input type="checkbox"/>	saut en parachute cambrai	 .fr (fr)	5
<input type="checkbox"/>	saut en parachute abbeville	 .fr (fr)	4
<input type="checkbox"/>	réservation saut en parachute vitry le françois	 .fr (fr)	4
<input type="checkbox"/>	réservation saut en parachute valenciennes	 .fr (fr)	2
<input type="checkbox"/>	réservation saut en parachute rethel	 .fr (fr)	4

Figure 5. Interface de l'outil de suivi de positionnement, MyPoseo

Pour vérifier que le contenu des balises HTML respecte les règles des moteurs de recherche, le référenceur peut utiliser un outil d'audit. Ainsi, toutes les balises sont visibles d'un seul coup d'œil. Cette démarche permet de gagner beaucoup de temps, car le référenceur n'est ainsi pas obligé de réaliser une vérification manuelle du code source pour chacune des pages.

- **Audit HTML** : notre entreprise pilote utilise l'outil *ScreamingFrog*²⁰, comme illustré sur la figure 6 ci-dessous. La première colonne affiche l'adresse URL de la page, la troisième colonne affiche le contenu de la balise <title> et la quatrième colonne affiche sa longueur.

²⁰ D'autres outils d'audit HTML existent comme Oncrawl, Botify Analytics ou SEO Quake

Chapitre 1. Le référencement

Internal	External	Protocol	Response Codes	URI	Page Titles	Meta Description	Meta Keywords	H1	H2	Images	Directives	AJAX	Custom	Analytics	Search Console
Filter: All Export															
				Address	Occurrences	Title 1		Title 1 Length							
1				http://www.esta-groupe.fr/	1	Ecole Supérieure de Technologie et des Affaires Belfort et Lyon » ESTA		70							
2				http://www.esta-groupe.fr/actualites/salons/	1	Salons étudiants ingénieur d'affaires, salons étudiants Lyon Paris » ESTA		73							
3				http://www.esta-groupe.fr/lecole/presentation-de-lesta/	1	Ecole d'ingénieurs d'affaires, école technico-commercial » ESTA		63							
4				http://www.esta-groupe.fr/contenu-esta/esta-belfort/	1	Forum Entreprises de l'ESTA Belfort le 29/09/2016 » ESTA		56							
5				http://www.esta-groupe.fr/recherche/presentation/	1	Projet de recherche en ingénierie d'affaires Lyon et Belfort » ESTA		67							
6				http://www.esta-groupe.fr/formation/filiere-chimie-biotech/	1	Formation d'ingénieur d'affaires Filière Chimie-Biotech » ESTA		62							
7				http://www.esta-groupe.fr/on-vous-rappelle/	1	On vous rappelle » ESTA		23							
8				http://www.esta-groupe.fr/etudiant/admission/	1	Concours école ingénieur d'affaires Lyon et Belfort » ESTA		58							
9				http://www.esta-groupe.fr/contact/contact	1	Contactez-nous » ESTA		21							
10				http://www.esta-groupe.fr/actualites/presse/	1	L'école ingénieur d'affaires Lyon et Belfort dans la presse « ESTA » ESTA		73							
11				http://www.esta-groupe.fr/lecole/les-campus/	1	Les campus étudiants technico-commercial Lyon et Belfort » ESTA		63							
12				http://www.esta-groupe.fr/lecole/esta-30-ans/	1	Ecole d'ingénieurs d'affaires, école technico-commercial » ESTA		63							
13				http://www.esta-groupe.fr/entreprise/apprentissage/	1	Recruter un stagiaire ingénieur commercial, technico-commercial » ESTA		70							
14				http://www.esta-groupe.fr/contact/contact/	1	Contactez-nous » ESTA		21							
15				http://www.esta-groupe.fr/entreprise/business-camp/	1	Business Camp » ESTA		20							
16				http://www.esta-groupe.fr/formation/les-stages/20-mois-dimmersion-en-entreprise/	1	20 mois d'immersion en entreprise » ESTA		40							
17				http://www.esta-groupe.fr/recherche/les-projets/	1	Projets de recherche école ingénieur d'affaires » ESTA		54							
18				http://www.esta-groupe.fr/contenu-esta/inscription-business-camp	1	Inscription aux business Camp à l'école d'ingénieur d'affaires » ESTA		69							
19				http://www.esta-groupe.fr/contenu-esta/telechargement-de-document/?file=Guide-Appre...	1	Télécharger la présentation de l'école d'ingénieur d'affaires » ESTA		68							
20				http://www.esta-groupe.fr/contenu-esta/stage/le-stage-de-1ere-annee/	1	Stage ouvrier 1ère année d'ingénieur d'affaires, école de commerce » ESTA		73							
21				http://www.esta-groupe.fr/recherche/lequipe-recherche/	1	Fiches Enseignants chercheurs à l'école d'ingénieur d'affaires » ESTA		69							
22				http://www.esta-groupe.fr/contenu-esta/stage/le-stage-de-3eme-annee/	1	Stage international de 3ème année d'ingénieur d'affaires, école de commerce » ESTA		82							
23				http://www.esta-groupe.fr/recherche/lequipe-recherche/thomas-rohr-2/	1	Enseignant-chercheur Thomas RÖHR » ESTA		39							
24				http://www.esta-groupe.fr/fiches-chercheurs-esta/dr-alain-gerome-fougeres/	1	Chercheur en modélisation dans les systèmes d'information coopératifs » ESTA		75							
25				http://www.esta-groupe.fr/fiches-chercheurs-esta/fiche-thomas-roehr/	1	Chercheur à l'école d'ingénierie en gestion de trafic urbain » ESTA		67							

Figure 6. Interface de l'outil d'audit SEO, affichage des balises <title>, ScreamingFrog

Il est important de connaître la quantité et la qualité des liens retour d'un site web pour mettre en place une stratégie de *netlinking*. Les outils d'analyse de liens retour permettent de faire l'inventaire de l'ensemble des pages web pointant vers un site web. Par ailleurs, ces outils permettent également d'horodater chaque lien. Il est donc possible pour un référenceur de recenser l'ensemble des liens retour obtenus par un site, de connaître les dates d'obtentions et de juger de leur qualité.

- **Analyse de liens retour :** notre entreprise utilise l'outil *Ahrefs*²¹, présenté sur la figure 7 ci-dessous. Le site web étudié dans notre exemple possède 1700 backlinks et 378 noms de domaines référents. En effet, un nom de domaine peut renvoyer plusieurs backlinks.

²¹ D'autres outils d'analyse de liens retour existent comme Majestic SEO ou Backlink Watch

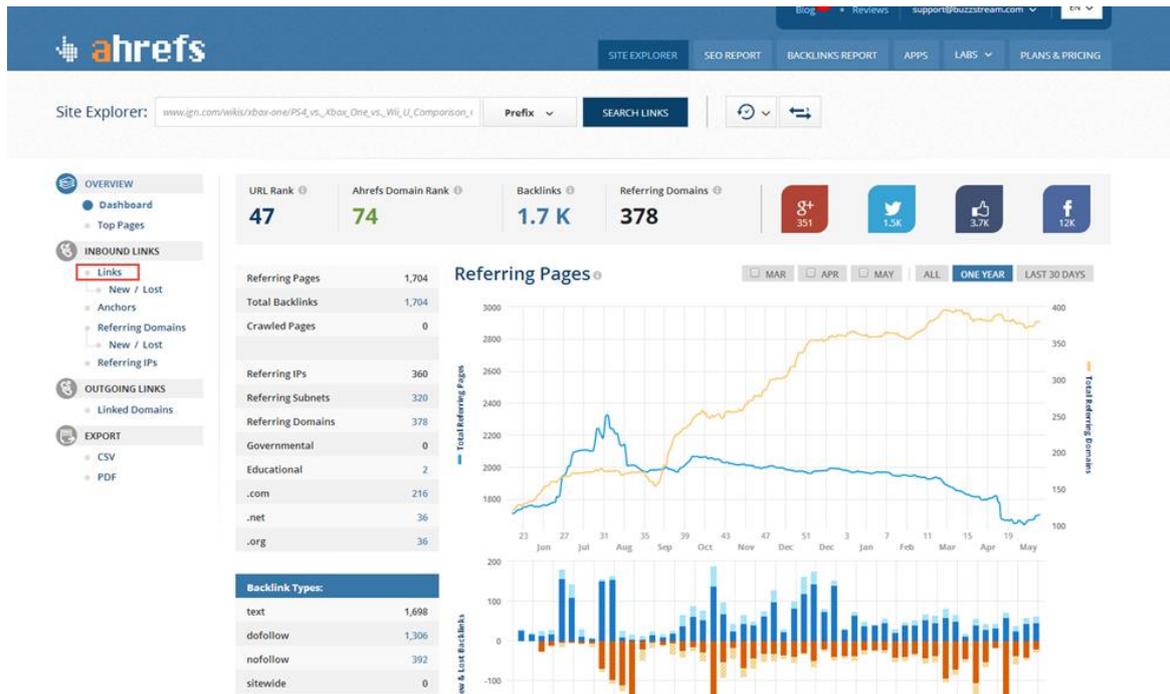


Figure 7. Interface de l'outil d'analyse de liens retour, Ahrefs

Le référencement doit être capable d'évaluer le potentiel d'un mot-clé pour espérer obtenir de bons résultats. Les outils d'analyse de mots-clés permettent de connaître le volume de recherche d'un mot-clé ainsi que la concurrence de celui-ci sur les moteurs de recherche.

- **Analyse de mots-clés :** notre entreprise pilote utilise l'outil *SemRush*²², présenté sur la figure 8 qui suit. Dans cet exemple, la concurrence du mot-clé « circus tickets » est de 0.83 pour un volume de recherche de 1000 requêtes par mois tandis que le nombre de résultats pour ce mot-clé est de 48 000 000.

²² D'autres outils d'analyse de mots-clés existent comme Yooda, SEMVisu ou KeywordSpy

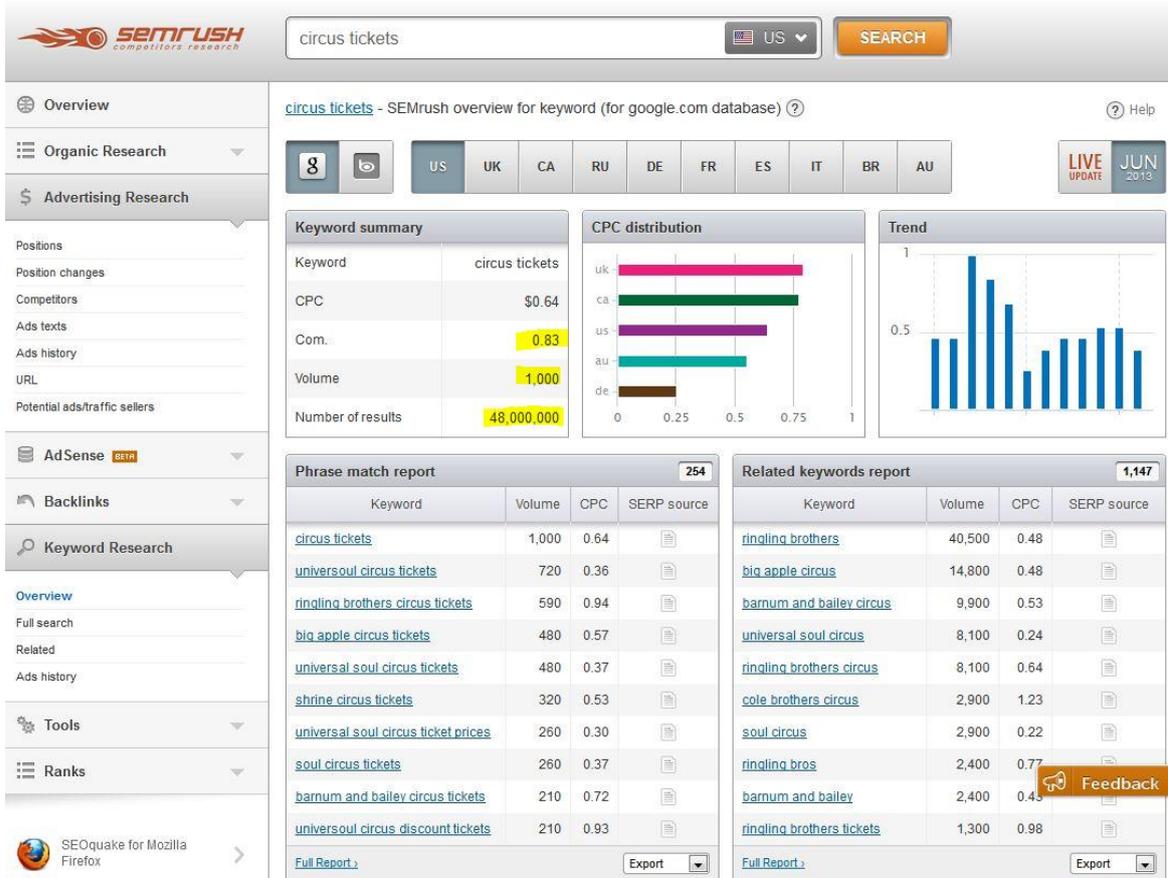


Figure 8. Interface de l'outil d'analyse de mots-clés, SemRush

Un critère déterminant pour le référencement est l'ancienneté du nom de domaine. Un web site d'une dizaine d'années remontera beaucoup plus rapidement sur une requête (même concurrentielle) qu'un site web créé il y a une semaine. Le référencement doit avoir connaissance de ce paramètre avant de s'engager sur des résultats attendus auprès de son client.

- **Analyse d'ancienneté :** pour connaître l'ancienneté d'un nom de domaine, les référenceurs de notre entreprise pilote utilisent le service « Whois »²³ fourni par les registres de nom de do-

²³ Contraction de l'anglais who is?, signifiant « qui est ? »

maine. Sur la figure 9 ci-dessous nous pouvons voir que le nom de domaine « utbm.fr » a été créé en 1995.



Whois Record for UtBm.fr	
- Whois & Quick Stats	
Dates	Created on 1995-09-29 - Updated on 2016-06-26
IP Address	193.48.231.29 - 23 other sites hosted on this server
IP Location	 - Franche-comte - Belfort - Institut Polytechnique De Sevenans

Figure 9. Interface du service Whois pour le nom de domaine utbm.fr, whois.domaintools.com

Dans cette section nous avons vu que le référenceur utilisait plusieurs outils et services lors de l'exercice de son activité. Que ce soit pour réaliser un reporting, analyser les critères *on-page*, pour analyser les critères *off-page*, analyser le potentiel des mots-clés ou pour consulter l'ancienneté du nom de domaine, le référenceur possède un support de logiciels et services conséquent. Le large choix d'outils proposés aux référenceurs peut être vécu comme un avantage, mais aussi comme un inconvénient, car il devra rester méthodique pour éviter de se noyer dans ce flot d'informations. Un système d'aide au référencement unique permettrait de pallier ce problème.

6. Discussion et conclusion

Dans ce chapitre nous avons présenté un état de l'art du référencement. Comme nous l'avons vu, les moteurs de recherche se doivent de fournir des résultats fiables et pertinents à leurs utilisateurs pour espérer conserver leurs parts de marché. Pour cela, ils doivent nécessairement faire **évoluer leurs algorithmes** de classement. Le référencement regroupe l'ensemble des techniques permettant d'améliorer la position d'un site web dans les pages de résultats des moteurs de recherche. À travers l'optimisation de critères *on-page* et *off-page*, le référenceur doit être capable d'améliorer la position du site web tout en utilisant des techniques autorisées par les moteurs, afin d'éviter des pénalités. Pour cela il pourra compter sur de nombreux outils et services à sa disposition pour exercer son activité. Encore faut-il les utiliser avec méthode et prendre du recul par rapport aux données récoltées.

Le référencement est une activité qui exige de **trouver des compromis** pour mettre en adéquation le besoin du client et le choix des actions techniques à implémenter sur le site. Pour définir les actions adéquates, le référenceur devra prendre en compte les futures mises à jour des algorithmes de classement des moteurs de recherche. Dès lors, un **travail de veille technologique** concernant les moteurs de recherche et de **prise d'information** fréquente concernant les critères est primordial pour être en capacité d'adapter ses techniques et ses décisions aux futurs changements.

Les **interactions avec le client** sont aussi importantes, car elles permettent d'identifier précisément les attentes et les besoins du client pour y répondre de manière adéquate et spécifique.

Par ailleurs, le panel d'outils à disposition du référenceur est large, ce qui demande une certaine rigueur pour traiter l'ensemble des données recueillies. Nous pensons que **regrouper ces outils** au sein d'une même interface pourrait faire gagner du temps et de l'efficacité au référenceur.

Nous verrons dans le prochain chapitre que le référencement pose également d'autres problèmes, le processus n'est pas formalisé ce qui entraîne des difficultés pour le maîtriser. Par ailleurs, l'incertitude des résultats et le nombre d'interactions en font un problème complexe. Dès lors, nous exposerons et nous détaillerons ces différents problèmes pour définir notre problématique de recherche et nos hypothèses de travail.

Chapitre 2. Problématique

1.	<i>Introduction</i>	38
2.	<i>Un problème complexe</i>	39
2.1	Structure dynamique et évolutive.....	39
2.2	Volume et diversité des données.....	41
2.3	Volume et diversité des interactions.....	42
2.4	Résultats incertains.....	43
2.5	Bilan.....	44
3.	<i>Le processus de référencement web</i>	45
3.1	Le référencement web : état de la littérature.....	45
3.2	Absence de cadre.....	46
3.3	Le référencement web est un processus.....	46
4.	<i>Problématique et hypothèses de travail</i>	47
5.	<i>Discussion et conclusion</i>	49

1. Introduction

Dans le chapitre précédent, nous avons défini le principe de référencement ainsi que l'ensemble des techniques permettant d'améliorer la position d'un site web sur les moteurs de recherche.

Dans ce chapitre nous proposons de présenter les problèmes qui peuvent intervenir lors de la pratique du référencement. Nous découvrirons notamment que le métier de référenceur reste en réalité très **artisanal**, car il repose en grande partie sur l'expérience vécue par les référenceurs avec leurs propres clients. Nous verrons également que le référencement peut être considéré comme un « **processus** » et non pas comme une « combinaison de techniques » comme suggéré souvent dans les travaux de littérature. Enfin, nous considérons le référencement comme un **problème complexe**, car il n'est pas clairement défini, les résultats sont incertains et longs à obtenir, et sa maîtrise est délicate.

L'objectif de ce chapitre est donc **d'exposer la problématique de recherche**, nos **hypothèses de travail** et de **proposer des méthodes et des outils** permettant d'y répondre.

La première section de ce chapitre expose la complexité de l'activité de référencement et récapitule les nombreux problèmes auxquels le référenceur peut être confronté pour l'exercice de son activité. La deuxième section présente comment le référencement est perçu dans la littérature et met en évidence l'absence de véritable processus de référencement. La troisième section expose notre problématique de recherche et nos hypothèses de travail. Nous terminerons ce chapitre par une discussion et une conclusion.

2. Un problème complexe

Pour l'exercice de son activité, le référenceur doit faire face à de nombreux problèmes. Cette section récapitule l'ensemble des problèmes inhérents à la maîtrise de l'activité de référencement, à savoir une structure dynamique et évolutive, le volume et la diversité des données et interactions, et enfin l'incertitude des résultats obtenus.

2.1 Structure dynamique et évolutive

Comme nous l'avons décrit au chapitre 1, le référencement est une activité mouvante, en constante évolution. Elle est étroitement liée aux mises à jour des algorithmes de classement des moteurs de recherche. Ces mises à jour sont perpétuelles, dès lors, le référenceur doit toujours garder à l'esprit que ses actions peuvent être remises en question à tout moment. Ainsi, des critères efficaces à l'instant t peuvent être inefficaces à l'instant $t + 1$.

Pour illustrer notre propos, nous pouvons prendre l'exemple du moteur de recherche Google. En 2011, le moteur a lancé un filtre nommé *Panda*²⁴ dont l'objectif était de nettoyer son index pour lutter contre les contenus de mauvaise qualité. 6% à 9% des requêtes en France ont été impactées par ce filtre et 12% pour les États-Unis²⁵. Les sites web avec des contenus de bonne qualité ont été récompensés par ce filtre

²⁴ <http://www.thesempost.com/understanding-google-panda-definitive-algo-guide-for-seos/>

²⁵ <https://webmasters.googleblog.com/2011/08/high-quality-sites-algorithm-launched.html>

et ont vu leur classement s'améliorer, à l'inverse les sites web avec du contenu de mauvaise qualité ont été sanctionnés. Les figures 10 et 11 ci-dessous illustrent l'impact positif et négatif du filtre *Panda* sur certains sites web français. Les trois sites les plus avantageés par le filtre Google *Panda* ont été *Amazon.com*, *Priceminister.com* et la *Fnac.com* avec respectivement +19,1%, +17,9% et +14,6% de visibilité en plus.

Les sites qui bénéficient le plus de Google Panda. Source: www.ranking-metrics.fr	Evolution de la visibilité en % (12/08/2011 par rapport au 24/07/2011)	RMX Informatique au 12/08/2011
amazon.fr	19,1%	3,72
priceminister.com	17,9%	2,63
fnac.com	14,6% source: www.ranking-metrics.fr	4,54
wikipedia.org	14,5%	1,98
clubic.com	13,8%	1,55
cdiscout.com	13,0%	4,23
ldlc.com	12,7%	3,19
ruecommerce.fr	12,2%	1

Figure 10. Sites web récompensés par le filtre Google Panda ainsi qu'en témoigne l'évolution positive de leurs visibilités, 2011, raking-metrics.fr

À l'inverse, les sites les plus désavantagés par ce filtre ont été *Twenga.fr*, *Wikio.fr* et *Webmarchand.com* qui ont respectivement perdu -54,2%, -51,8% et -50,6% de visibilité.

Les sites les plus pénalisés par Google Panda. Source: www.ranking-metrics.fr	Evolution de la visibilité en % (12/08/2011 par rapport au 24/07/2011)	RMX Informatique 12/08/2011
twenga.fr	-54,20%	10,82
wikio.fr	-51,80%	29,1
webmarchand.com	-50,60%	44,23
achetezfacile.com	-50,4%	32,69
ciao.fr	-49,60% source: www.ranking-metrics.fr	8,54
presence-pc.com	-49,00%	9,64
infos-du-net.com	-45,60%	10,47
touslesprix.com	-45,50%	29,74
leguide.com	-43,00%	4,02
shopping.com	-38,40%	17,27
mobiles-actus.com	-34,70%	35,23
acheter-moins-cher.com	-34,30%	15,15

Figure 11. Sites web pénalisés par le filtre Google Panda ainsi qu'en témoigne l'évolution négative de leurs visibilités, 2011, raking-metrics.fr

Cet épisode ne fut pas un cas unique puisqu'en 2012, Google a récidivé en lançant le filtre *Penguin* qui a pour objectif de sanctionner les sites web possédant une grande quantité de liens retour (backlinks) de mauvaise qualité. Cette fois-ci, environ 3% des requêtes dans le

monde ont été impactées²⁶. Comme avec *Panda*, certains sites ont gagné tandis que d'autres ont perdu en visibilité.

Les fondements sur lesquels repose l'activité de référencement évoluent de façon dynamique. Ils peuvent être remis en question à tout moment. Cela découle d'une connaissance limitée des algorithmes de classement des moteurs de recherche. Mais le référenceur doit également rester vigilant, car les premières positions sur les moteurs de recherche sont très convoitées. C'est la raison pour laquelle il doit également prendre en compte le travail des autres référenceurs. En effet, sur une requête très concurrentielle, il n'est pas rare que tous les sites web de la première page adoptent une stratégie de référencement. Pour maîtriser l'activité de référencement, il faudra donc prendre en compte sa **structure dynamique et évolutive**.

Nous retiendrons que l'environnement de travail du référenceur est en constante évolution ce qui exige une grande capacité d'adaptation. En effet, le référenceur doit analyser constamment l'impact de ses actions pour être en capacité de réagir rapidement. Cette démarche perpétuelle permet de diminuer le risque d'être déclassé dès l'arrivée d'un nouveau concurrent ou lors d'une prochaine mise à jour du moteur de recherche.

En conséquence, nous considérons que : « *Le référencement est une activité continue* ».

2.2 Volume et diversité des données

L'activité de référencement exige également le traitement d'un grand nombre de données que constituent les critères. En effet, si au commencement du web seulement quelques critères étaient utilisés par les moteurs de recherches, ils utilisent aujourd'hui des dizaines, voire des centaines de critères pour déterminer le classement d'un site

²⁶ <https://search.googleblog.com/2012/04/another-step-to-reward-high-quality.html>

web²⁷. Dès lors, le référenceur doit être en mesure de **se tenir informé** sur les critères ayant le plus d'impact en réalisant une **veille technologique**. Cela lui permet d'être encore plus efficace et de concentrer ainsi son attention sur l'essentiel afin de ne pas perdre de temps dans ses activités de référencement. En effet, le référenceur doit respecter des **impératifs de temps et de coût**.

En conséquence, nous considérons que : « *L'acquisition de données sur l'environnement permet d'améliorer l'efficacité du référencement* ».

2.3 Volume et diversité des interactions

Durant un projet de référencement, des interactions apparaissent à différents niveaux. Le premier type d'interactions a lieu entre les acteurs d'un projet que sont le client et le référenceur. Durant un projet de référencement, ils peuvent échanger à propos des besoins, des objectifs à atteindre et des délais à respecter, mais aussi des contraintes et des problèmes rencontrés. Nous verrons dans le chapitre suivant que différents niveaux d'interactions émergent selon le niveau de connaissance et l'expérience du client. D'autres types d'interactions apparaissent, notamment entre le référenceur et ses outils. C'est notamment le cas quand le référenceur soumet un mot-clé à un outil de positionnement et que celui-ci lui retourne une position donnée. Ces **interactions sont nombreuses et hétérogènes**, elles peuvent faire perdre du temps et de l'efficacité au référenceur. C'est la raison pour laquelle elles doivent être maîtrisées et centralisées afin de **privilégier leur qualité** [36] plutôt que leur volume.

En conséquence, nous considérons que : « *Des interactions de qualité entre acteurs et outils permettent d'améliorer les résultats du référencement* ».

²⁷ Le moteur de recherche Google compte plus de 200 critères de classement, <https://support.google.com/webmasters/answer/70897?hl=fr>

2.4 Résultats incertains

L'activité de référencement a la particularité d'aboutir à des résultats incertains. Si dans la majorité des cas les référenceurs parviennent à améliorer la position d'un site web sur les moteurs de recherche, ce n'est pas toujours systématique. En effet, les positions peuvent stagner voir régresser [37]. Les trois courbes sur la figure 12 ci-dessous exposent l'évolution de la position de trois sites web différents après l'optimisation des balises <title> :

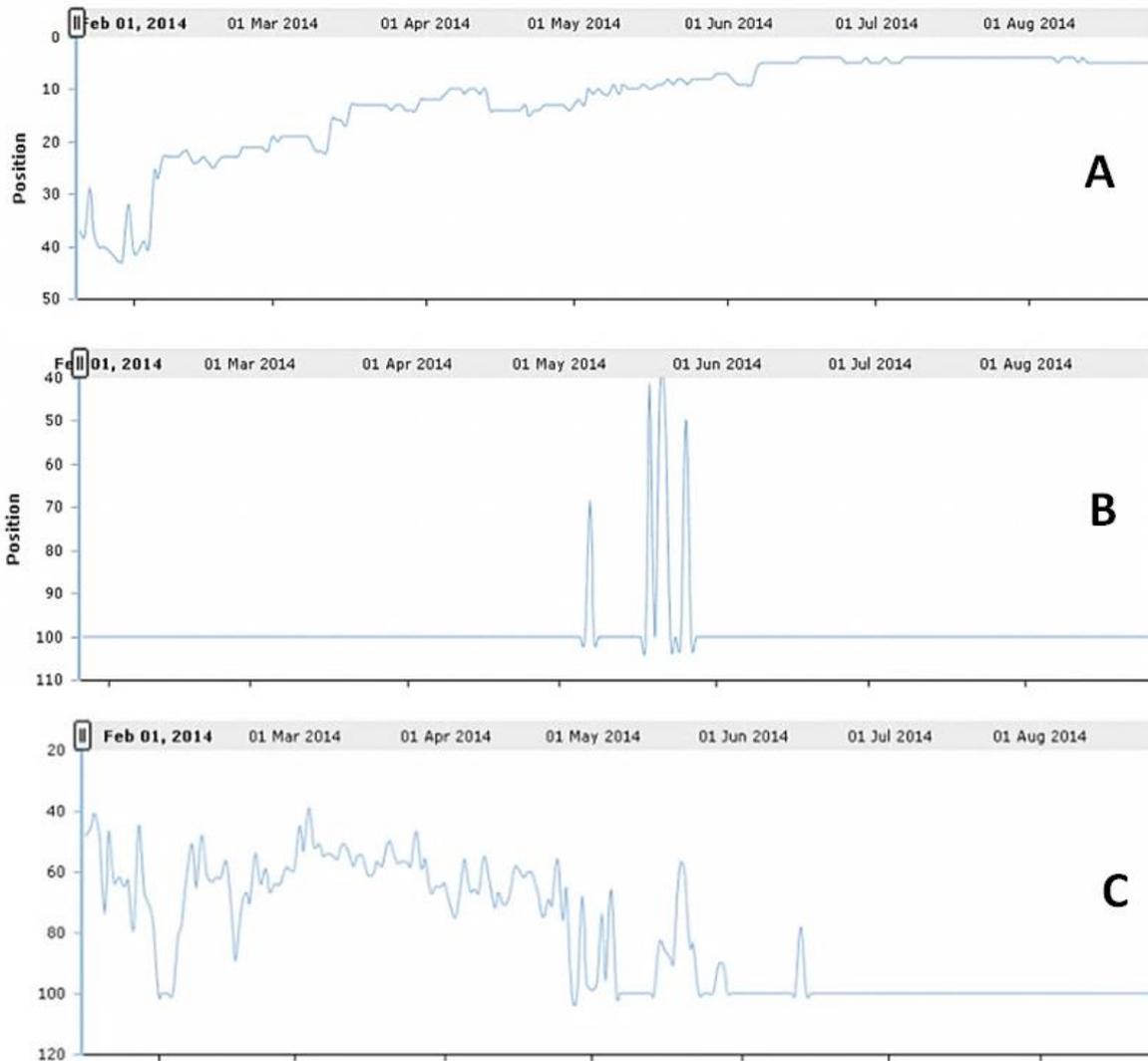


Figure 12. Trois types d'évolution de classement possibles : (A) augmentation, (B) stagnation et (C) régression

Comme nous pouvons le remarquer, malgré la mise en place d'une technique d'optimisation identique, l'évolution du classement est in-

certaine. En effet, la courbe (A) expose une augmentation du classement, la courbe (B) expose une stagnation du classement et la courbe (C) expose une régression du classement.

Par ailleurs, les résultats ne sont pas immédiats. Les actions de référencement peuvent mettre d'une semaine à quelques mois avant de donner des résultats. Les délais de prise en compte dépendent grandement de l'ancienneté du nom de domaine et de la concurrence sur le terme visé. Si le référenceur peut s'engager aisément sur une hausse de position, il lui sera difficile de la quantifier. Le référenceur doit donc faire face à cette incertitude avant d'annoncer les résultats attendus à son client.

En conséquence, nous considérons que : « *Le référencement a des résultats incertains* ».

2.5 Bilan

Un problème peut être considéré comme simple « si l'ensemble des éléments nécessaires à sa résolution sont connus ou accessibles. Même s'il s'agit d'une situation nouvelle, le problème peut être résolu par une procédure systématique ou l'application de règles déjà connues. La solution est unique, non équivoque et vérifiable. En contraste, dans un problème complexe, les informations peuvent être incomplètes, imprécises et contradictoires. Un problème complexe oblige à prendre en compte plusieurs aspects ou schémas de référence et différentes représentations du problème peuvent être définies selon les présupposés, informations et opinions considérés. Aucune procédure ou technique analytique spécifique ne permet d'arriver à une solution unique et non équivoque. Ainsi, plusieurs solutions sont possibles et des critères sont à déterminer afin de juger la pertinence et la valeur d'une solution à la lumière de notre connaissance actuelle de la question »²⁸. En raison de sa structure dynamique et évolutive, de la connaissance limitée des méthodes et des critères de classement, du

²⁸ <http://app.cegep-ste-foy.qc.ca/index.php?id=612>

volume et de la diversité des données, du nombre important d'interactions et de l'incertitude des résultats, nous considérons le référencement comme **un problème complexe**.

En conséquence, nous considérons que : « *Le référencement est un problème complexe* ».

3. Le processus de référencement web

3.1 Le référencement web : état de la littérature

Dans la section précédente, nous avons mis en évidence les problèmes liés à la pratique du référencement et notamment le fait que les résultats peuvent être incertains, car les moteurs de recherche changent constamment leurs méthodes de classement. Malaga [38] s'est d'ailleurs interrogé sur la question : « *Comment un propriétaire de site web peut-il planifier et gérer une quelconque stratégie dans une situation aussi volatile ? Existe-t-il des techniques ou des méthodes étant susceptibles de fournir des résultats sur le long terme ?* ». Pour reprendre les termes de Malaga, selon lui des « *techniques* » ou des « *méthodes* » pourraient aider le référenceur à faire perdurer ses actions.

L'orientation choisie par Malaga n'est pas isolée. Une étude que nous avons réalisée [37] auprès de cinquante articles scientifiques relatifs au domaine du référencement (naturel), dont les travaux fondateurs de Evans [21] et Zhang et Dimitroff [22], [23], nous permet de mettre en évidence que les chercheurs de manière générale ne considèrent pas l'aspect « évolutif » du référencement. En effet, la grande majorité de ces articles se contentent d'exposer les critères à optimiser (souvent sous forme de listes ou de guides). Mis à part Malaga [38], aucun chercheur à notre connaissance ne se questionne sur le fait que les mises à jour des algorithmes de classement peuvent reconsidérer l'efficacité du travail de référencement.

Comme nous l'avons découvert [37], l'état de la littérature considère plutôt le référencement comme la mise en place de combinaisons de « techniques », et si possible de « techniques autorisées » par les moteurs de recherche. Comme nous le verrons dans les sections suivantes, notre démarche se distingue de celle décrite dans la littérature

existante, car nous considérons en effet le référencement comme un processus et non pas comme une combinaison de techniques. Selon la norme ISO 9000 [39], un processus est : « *un ensemble d'activités transformant des éléments d'entrée en éléments de sortie* ». Ainsi, pour le processus de référencement, les éléments d'entrée sont les pages web à référencer, et les éléments de sortie sont les pages web dont la position s'est améliorée grâce à l'action de référencement.

En conséquence, nous considérons que : « ***Le référencement est un processus*** ».

3.2 Absence de cadre

Depuis que la profession existe, les référenceurs ne disposent pas de véritable cadre à leur activité [4]. Nous avons pu le vérifier dans l'entreprise pilote où nos travaux de recherche ont été conduits. Si les étapes « d'analyse du besoin client » et de « suivi des résultats » sont correctement intégrées, celles concernant « l'optimisation des pages » se distinguent par leur **manque d'harmonisation**. Même si les référenceurs connaissent les critères à forts impacts de par leurs diverses expériences et projets réalisés, quand ils optimisent les pages d'un site web, ils sont seuls juges de leurs actions. Ainsi le référenceur n'obéit pas à un cadre précis et c'est à lui d'évaluer la nécessité de modifier tels ou tels critères pour améliorer le classement des pages. En réalité si l'activité de référencement est aujourd'hui pérenne, elle n'en demeure pas moins une **activité artisanale**.

Comme nous le verrons dans la section suivante, la création d'un processus peut permettre de mieux cadrer cette activité. Par ailleurs, la conception d'un système pour aider les référenceurs à mieux structurer leur activité est aussi l'un des axes de nos recherches comme nous le verrons dans le chapitre 5.

3.3 Le référencement web est un processus

A la fin de la section 2, nous avons proposé de considérer le référencement comme un **problème complexe**. Sa résolution implique de la rigueur et les prises de décision doivent être orientées dans un cadre précis. Les travaux de la littérature sont focalisés sur la façon de « **combinaison des techniques** » dans le but d'obtenir des solutions effi-

caces et ne considèrent pas son « **aspect évolutif** ». Pour nous, le travail d'un référenceur et les actions qu'il doit mener ne doivent pas se limiter à une simple combinaison de techniques. Il doit identifier et comprendre le problème de départ et à travers les informations collectées, il doit aussi pouvoir apporter des solutions adaptées. Pour rappel, la **conception et le développement** se définissent comme « *une somme d'activités créatrices qui, partant des besoins exprimés et des connaissances existantes, aboutissent à la définition d'un produit satisfaisant ces besoins et industriellement réalisables* » [40]. C'est la raison pour laquelle nous avons choisi de considérer l'activité de référencement comme un « **processus** », et plus particulièrement comme un « **processus de conception** ». En effet, comme pour un processus de conception classique [41]–[44], le processus de référencement intègre les étapes suivantes :

- 1) **Définition du problème** (ex : Quel gain de position est espéré ? Sur quels mots-clés ?)
- 2) **Collecte des informations** (ex : nombre de pages, technologie web utilisée par le site, compatibilité du site avec les techniques de référencement)
- 3) **Génération de solutions** (ex : corriger les balises <title>, ajouter des mots-clés dans le texte)
- 4) **Analyse et sélection des solutions** (ex : les balises <title> ne sont pas modifiables sur le site, la solution retenue sera l'ajout de mots-clés dans le texte)
- 5) **Test et implémentation de la solution** (ex : ajout des mots-clés, suivi du gain de position)

Dans la section suivante, nous présenterons notre problématique de recherche ainsi que nos hypothèses de travail.

4. Problématique et hypothèses de travail

Nous avons vu qu'en raison de sa structure dynamique et évolutive, de la connaissance limitée des méthodes et des critères de classement, du volume et de la diversité des données, du nombre important d'interactions et de l'incertitude des résultats, nous proposons de considérer le référencement comme un **problème complexe**. Par ailleurs, nous estimons que l'activité de référencement n'est pas une

combinaison de techniques comme le prétend la littérature, mais un processus, et plus particulièrement un **processus de conception**. Mais ce processus n'est pas défini ce qui rend sa conception et sa maîtrise délicate d'autant que l'activité de référencement est très **artisanale**. Nous pouvons alors formuler la problématique de recherche suivante :

Problématique : « *Comment concevoir le processus de référencement web pour améliorer sa maîtrise par le couple référencieur/client ?* »

Le référencement est une activité artisanale qui peut être structurée, dès lors nous pouvons créer et définir un processus. Notre première hypothèse est donc la suivante :

Hypothèse n°1 : « *Le processus de référencement est un processus complexe et structurable* »

Après avoir structuré l'activité de référencement à travers la définition d'un processus, nous avons ensuite souhaité le modéliser. Pour ce faire, nous avons utilisé le paradigme agent qui selon nous propose le meilleur compromis pour modéliser le processus de référencement. Notre deuxième hypothèse est donc la suivante :

Hypothèse n°2 : « *Le processus de référencement est modélisable* »

Enfin, à partir du modèle que nous aurons établi, nous souhaitons améliorer la maîtrise du processus de référencement. Pour ce faire, nous souhaitons développer un système d'aide à la décision servant de support à l'activité de référencement, et notamment au couple référencieur/client. La troisième hypothèse est donc la suivante :

Hypothèse n°3 : « *Le processus de référencement est mieux maîtrisable grâce à l'utilisation d'un système d'aide à la décision* »

En répondant à ces trois hypothèses de travail, nous pensons pouvoir répondre favorablement à notre problématique de recherche.

5. Discussion et conclusion

Dans ce chapitre nous avons mis en évidence plusieurs problèmes qui caractérisent l'activité de référencement. Notre première hypothèse de travail nous permettra de vérifier si le processus de référencement, que nous considérons comme complexe, peut être **structuré**. Puis nous tenterons de **modéliser** ce processus ; nous répondrons ainsi à notre deuxième hypothèse. Enfin, le développement d'un système d'aide à la décision conçu à partir du modèle que nous aurons établi permettra de **mieux maîtriser** le processus de référencement, répondant ainsi à notre troisième hypothèse. En apportant une réponse à notre problématique de recherche, nos travaux permettront au couple référenceur/client de mieux assimiler, mais aussi de mieux maîtriser le processus de référencement.

A notre connaissance, la littérature évoque très peu les problèmes liés référencement ; Malaga [38] est l'un des seuls à se questionner sur la durabilité des actions de référencement. En effet, la majorité des chercheurs ont concentré leur recherche sur l'identification des critères et des techniques pour améliorer le référencement [37]. Pourtant l'influence des critères et l'efficacité des techniques de référencement peuvent être remise en question à tout moment, ce qui oblige les référenceurs à rester vigilants pour espérer viser des objectifs sur le long terme.

L'aspect « processus de conception » à notre connaissance n'a pas été évoqué dans la littérature pour caractériser l'activité de référencement. Nous pensons que cet axe de recherche mérite d'être exploré, car considérer le référencement comme un processus permettra au référenceur d'identifier des solutions adéquates pour répondre au besoin du client. Cette approche est déjà utilisée pour la conception de produits/services [45], nous pensons qu'elle peut être adaptée pour répondre notre problème.

Dans le prochain chapitre, nous définirons le processus de référencement et ses différentes étapes à partir d'un méta-modèle d'ingénierie issu du domaine de la conception de produits/services. Cette démarche nous permettra de répondre à notre première hypothèse de travail.

Chapitre 3. Le processus de référencement web

1.	<i>Introduction</i>	50
2.	<i>Méthode pour définir le processus de référencement web</i>	51
2.1	Similitude avec la conception de produits/services.....	51
2.2	Utilisation d'un méta-modèle d'ingénierie	52
2.3	Identification du processus de référencement web	54
2.4	Bilan	56
3.	<i>Acteurs et éléments du processus de référencement web</i>	57
3.1	Les acteurs humains	57
3.1.1	Le référencieur.....	57
3.1.2	Le client	57
3.2	Les étapes du processus de référencement web.....	58
3.2.1	Étape 1 : Identification du besoin	58
3.2.2	Étape 2 : Étude de la faisabilité du projet.....	59
3.2.3	Étape 3 : Proposition des actions.....	59
3.2.4	Étape 4 : Implémentation des actions	60
3.2.5	Étape 5 : Mesure de l'impact.....	60
3.3	Interactions entre les acteurs du processus de référencement web.....	62
4.	<i>Collaboration et niveau de connaissance</i>	63
4.1	Deuxième version du processus de référencement web	65
5.	<i>Discussion et conclusion</i>	68

1. Introduction

Dans le chapitre précédent, nous avons mis en évidence les problèmes liés à l'activité de référencement et nous avons montré l'importance de considérer le référencement comme un processus et non pas comme une combinaison de techniques.

L'activité de référencement est artisanale, car le processus de référencement repose essentiellement sur des bases techniques acquises au fur et à mesure des projets réalisés. Malgré son caractère complexe, nous pensons que ce processus peut être structuré. Pour cela, il est nécessaire de le définir afin d'identifier les étapes clés et mettre en évidence les acteurs impliqués ainsi que leurs interactions.

L'objectif de ce chapitre est ainsi de répondre à notre première hypothèse :

Hypothèse n°1 : « *Le processus de référencement est un processus complexe et structurable* »

En répondant à cette hypothèse, il nous sera possible dans le prochain chapitre de modéliser ce processus.

Dans la première section de ce chapitre, nous proposons d'utiliser un méta-modèle d'ingénierie pour définir le processus de référencement. Dans la deuxième section, nous présentons les acteurs et les étapes de ce processus. Dans la troisième section, nous expliquons comment la collaboration ainsi que le niveau de connaissance du client permettent de définir d'autres versions du processus de référencement. Nous terminons par une discussion et une conclusion.

2. Méthode pour définir le processus de référencement web

2.1 Similitude avec la conception de produits/services

Dans notre étude nous considérons le référencement comme un processus et non pas comme une combinaison de techniques. En effet, le référencement entretient des interactions complexes avec son environnement. L'environnement est composé en partie par les algorithmes des moteurs de recherche, la technologie, les pages web, les clients, les référenceurs et d'autres facteurs qui influencent directement ce processus. Ainsi pour avoir une meilleure connaissance des interactions, des éléments d'entrée et de sortie impliqués dans ce processus, nous nous sommes demandé : « comment le concevoir ? »

Cette question est typiquement un problème de conception. Pour cela, nous avons choisi de nous inspirer du domaine de la conception de produits/services qui considère le produit/service comme étant « *le résultat à atteindre* » et le processus comme étant « *le chemin à parcourir pour atteindre le résultat* ». Un processus de conception sert donc à déterminer comment concevoir le meilleur produit/service [40], [46]. Le processus de référencement quant à lui, sert à déterminer les meilleures actions pour référencer un site web sur les mots-clés choisis. L'objectif est donc quasiment similaire, dans les deux cas l'objectif doit permettre de trouver des solutions viables permettant de satisfaire les besoins et les attentes du client tout en prenant en compte les contraintes de l'environnement. Dans la section sui-

vante, nous présentons notre méthode pour définir le processus de référencement.

2.2 Utilisation d'un méta-modèle d'ingénierie

Certains chercheurs [47] ont choisi d'utiliser des méta-modèles²⁹ pour élaborer des processus de conception de produits/services. Cette approche consiste à formaliser la démarche de conception en décomposant les différentes étapes du processus afin de les analyser de manière précise et indépendante. Nous proposons de réutiliser cette démarche pour définir le processus de référencement.

Pour cela, nous avons sélectionné le méta-modèle d'ingénierie proposé par Ostrosi et al. [47]. Il a déjà été utilisé pour modéliser des processus de conception de produits, mais sa portée générale lui permet d'être utilisé pour définir d'autres types de processus de conception.

Ce méta-modèle est composé de quatre « modèles » décrits ci-dessous :

- **Modèle conceptuel** : il est basé sur des concepts et des règles. C'est le premier aspect (besoins clients ou règles techniques) à comprendre avant la conception d'un produit ou d'un processus. Dans notre cas, cela peut être l'expression d'un concept technique du référencement (ex : « *la balise <title> doit être unique sur chaque page d'un site web* »). Ce modèle contient des concepts et des règles prédéfinis.
- **Modèle mathématique** : c'est un modèle concret. Il s'exprime d'un point de vue mathématique. Cela peut être une équation différentielle, une formulation logique, ou une structure mathématique. (ex : « *une formule mathématique qui fait le lien entre une technique de référencement et l'évolution du classement* »).

²⁹ Méta-modèle : littéralement modèle de modèle, il peut être défini comme la représentation d'un point de vue particulier sur des modèles, <http://www.ie.inf.uc3m.es/ggenova/Warsaw/Part3.pdf>

- **Modèle informatique** : c'est la représentation du modèle mathématique d'un point de vue informatique. Dans notre cas cela peut être un algorithme ou un outil permettant de faciliter le processus de référencement. (ex : « *un logiciel de génération de mots-clés, un outil de simulation, un outil de monitoring, etc.* »)
- **Modèle expérimental** : il est dédié à l'expérimentation. Les résultats sont des observations ou des tendances. Dans notre cas, cela peut être : « *une expérimentation réalisée sur des pages web pour tester l'influence d'un critère sur le classement* ».

Les quatre modèles présentés ci-dessus possèdent des définitions différentes, mais parfois les frontières entre ces différents modèles sont difficilement perceptibles. Par exemple, un modèle informatique ou une règle conceptuelle exprimée de manière très formelle peuvent très bien appartenir au modèle mathématique. Néanmoins, pour plus de clarté, les éléments présentés dans nos travaux de recherche appartiendront toujours à un seul et unique modèle.

Le meta-modèle est également composé de quatre « domaines » présentés ci-dessous :

- **Domaine client** : il exprime un besoin client. (ex : « *Je souhaite que mon site web atteigne la troisième position sur ce mot-clé* »). En comprenant le domaine client, il est possible de répondre de manière adéquate aux besoins du client.
- **Domaine fonctionnel** : la conception de processus doit être modélisée en utilisant des fonctions liées à la structure, à l'environnement et au comportement du processus. Des exemples de fonctions pourraient être : « *améliorer le trafic issu des moteurs de recherche vers le site web* » ou « *améliorer le classement du site web* ».
- **Domaine physique** : il correspond à des solutions qui répondent au domaine fonctionnel. (ex : « *remplir les balises* »).

<title> sur les différentes pages » ou « répéter des mots-clés pertinents dans le contenu texte »).

- **Domaine processus** : il s’agit des éléments de contraintes liées au domaine physique. (ex : « être compatible avec les mises à jour des moteurs de recherche », « être compatible avec la technologie web utilisée sur le site web »).

Le tableau 1 ci-dessous illustre le méta-modèle décrit :

Tableau 1. Méta-modèle d’ingénierie issu de la conception de produits, [47]

	Domaine client	Domaine fonctionnel	Domaine physique	Domaine processus
Modèle conceptuel	Modèle conceptuel du domaine client	Modèle conceptuel du domaine fonctionnel	Modèle conceptuel du domaine physique	Modèle conceptuel du domaine processus
Modèle mathématique	Modèle mathématique du domaine client	Modèle mathématique du domaine fonctionnel	Modèle mathématique du domaine physique	Modèle mathématique du domaine processus
Modèle informatique	Modèle informatique du domaine client	Modèle informatique du domaine fonctionnel	Modèle informatique du domaine physique	Modèle informatique du domaine processus
Modèle expérimental	Modèle expérimental du domaine client	Modèle expérimental du domaine fonctionnel	Modèle expérimental du domaine physique	Modèle expérimental du domaine processus

2.3 Identification du processus de référencement web

Pour proposer une première version de processus, nous avons conçu le référencement en utilisant une démarche de conception. Ainsi, en reprenant le méta-modèle [47] qui avait permis la validation d’un processus de configuration de produit [45], [48], nous avons pu isoler certaines étapes clés faisant partie de ce processus. Ainsi nous avons pu identifier une première version du processus de référencement comme le montre la figure 13 qui suit :

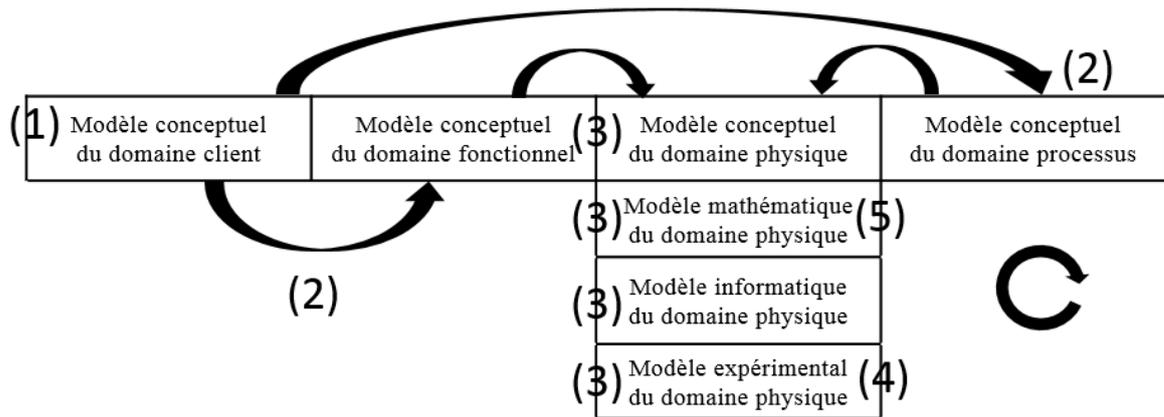


Figure 13. Première version du processus de référencement

Ce processus se décompose en cinq étapes que nous décrivons ci-dessous :

Étape 1 : Durant cette étape le client exprime son besoin en référencement. (ex : « *Je souhaite positionner mon site web à la troisième position sur ce mot-clé* »). Cette étape est purement **conceptuelle** et elle appartient uniquement au **domaine client**.

Étape 2 : Le besoin du client est ensuite traduit en *fonctions* par le référencier. Cela se fait suite à plusieurs itérations avec le client afin de comprendre précisément son besoin. Ensuite le référencier identifie des *contraintes* pour étudier la faisabilité du projet. Par exemple : « *être compatible avec les mises à jour des moteurs de recherche* », « *être compatible avec la technologie utilisée par le site web* ». À noter que le client ne participe pas à l'identification des *contraintes*. Durant cette étape **conceptuelle**, seul le **domaine fonctionnel** propre aux *fonctions* et le **domaine processus** propre aux *contraintes* sont concernés.

Étape 3 : Quand toutes les *fonctions* et les *contraintes* sont correctement définies, le référencier propose des *solutions* techniques. Elles peuvent appartenir aux quatre modèles (*Conceptuel*, exemple : *le référencier identifie une liste de concepts et règles techniques à implémenter*, *Mathématique*, exemple : *le référencier crée une équation combinant les besoins du client et les résultats qu'il souhaite obtenir*, *Informatique*, exemple : *le référencier adapte le code HTML du site*, *Expérimental*, exemple : *le référencier expose*

des résultats expérimentaux au client que celui-ci peut choisir d'appliquer sur son site). Durant cette étape, seul le **domaine physique** propre aux *solutions* est concerné. Les *solutions* quant à elles, peuvent appartenir à **tous les modèles**.

Étape 4 : Une fois que le client a validé les *solutions* proposées par le référenceur, celui-ci doit les implémenter dans le site web. Cette étape d'implémentation concerne toujours le **domaine physique** propre aux *solutions*, mais l'implémentation des actions touche uniquement le **modèle expérimental**.

Étape 5 : Une fois que les *solutions* ont été implémentées, le référenceur mesure l'impact de ses actions. Cette étape reste dans le **domaine physique** propre aux *solutions*, mais la mesure des résultats concerne uniquement le **modèle mathématique**.

Après avoir mesuré l'impact de ses actions et obtenu le retour du client, si les *solutions* n'apportent pas les résultats escomptés, le référenceur devra reconsidérer chaque étape : le processus de référencement recommence, il est **cyclique**.

2.4 Bilan

Le méta-modèle que nous avons utilisé nous a permis d'identifier une première version du processus de référencement. Celui-ci se décompose en cinq étapes que nous appellerons :

Étape 1 : Expression du besoin

Étape 2 : Étude de la faisabilité du projet

Étape 3 : Proposition des actions

Étape 4 : Mise en place des actions

Étape 5 : Mesure de l'impact

Lors du processus de référencement, l'objectif du référenceur est d'optimiser le nombre d'échanges avec le client tout en trouvant des solutions efficaces permettant d'améliorer en un minimum de temps le classement du site web. À travers cette décomposition du méta-modèle nous avons pu déterminer les principales étapes du processus de référencement et mettre en évidence son caractère **cyclique**. Pour être viable, le processus de référencement doit également être **dyna-**

mique afin de pouvoir rester adaptable et efficace à tout changement de contexte de travail et d'environnement.

Pour continuer sur cet axe et définir précisément le processus de référencement, il est désormais nécessaire de décrire en détail les étapes du processus de référencement, les acteurs concernés par ce processus, ainsi que leurs interactions. Par ailleurs nous nous sommes demandé si le processus de référencement était collaboratif [5]. Nous allons apporter des réponses à cette question dans la section suivante.

3. Acteurs et éléments du processus de référencement web

3.1 Les acteurs humains

Comme nous l'avons constaté dans notre entreprise pilote, les acteurs humains concernés par une activité de référencement sont généralement au nombre de deux : le client et le référenceur. Néanmoins, il est possible que plusieurs interlocuteurs soient impliqués du côté du client (ex : chef d'entreprise, responsable informatique, responsable marketing, etc.). Le référenceur de son côté peut aussi être assisté par des collègues (ex : développeur, rédacteur web, graphiste, etc.). Cependant, pour simplifier l'élaboration du processus de référencement, nous limitons volontairement le cercle des acteurs humains au couple référenceur/client.

3.1.1 *Le référenceur*

Le référenceur est un acteur humain faisant partie intégrante du processus de référencement, comme vu au chapitre 1 il possède l'expertise métier de par sa formation et son expérience. Un référenceur est en relation avec un client. Ils collaborent ensemble lors d'un « projet de référencement » qui les lie contractuellement. Le référenceur réalise plusieurs tâches, il repère les lacunes d'un site web, il propose des techniques adaptées permettant d'améliorer le classement du site et enfin il assure le suivi des résultats, toujours en accord avec les besoins du client.

3.1.2 *Le client*

Le client de son côté est également un acteur humain faisant partie du processus de référencement. C'est lui qui exprime le besoin, c'est

également le propriétaire du site web dont il possède les accès permettant de modifier sa structure et son contenu. Dans la section 4, nous montrerons que le client peut même collaborer directement avec le référenceur lors du processus de référencement, en fonction de son niveau de connaissance.

3.2 Les étapes du processus de référencement web

La section précédente nous a permis de présenter les acteurs du processus de référencement. L'élaboration de notre processus nécessite également de décrire précisément les étapes nécessaires à son déroulement. Grâce à des observations réalisées dans notre entreprise pilote, et après avoir interrogé plusieurs référenceurs concernant leurs méthodes de travail, nous avons pu mettre en place une méthodologie permettant de répertorier cinq étapes. Nous présentons chacune d'elles ci-dessous.

3.2.1 Étape 1 : Identification du besoin

Le processus de référencement débute toujours par l'identification du besoin. Durant cette étape, le client communique une liste de mots-clés au référenceur pour lesquels il souhaite améliorer son classement. Cette liste de mots-clés est stratégique pour la réussite du projet. Les délais pour obtenir des résultats peuvent être considérablement réduits ou augmentés en fonction du choix des mots-clés. En effet, si les mots-clés choisis sont trop génériques (forte concurrence) et ne sont pas en accord avec la thématique du site web, les résultats peuvent être décevants. C'est la raison pour laquelle le référenceur peut dans certains cas, proposer au client de modifier cette liste. Une fois que le client a choisi les mots-clés qu'il souhaitait conserver, le référenceur génère le premier rapport de positionnement. Ce rapport appelé "*rapport point zéro*" dans notre entreprise pilote dont un exemple est présenté sur la figure 14, permet de consigner la position initiale du site web sur les différents mots-clés avant le lancement du projet de référencement. Par exemple, pour le mot-clé « *joint de dilatation* », la position du site web au 24 juin 2016 était de 59 soit la 6^e page sur le moteur de recherche Google.fr, mais cette position n'est pas définitive. L'évolution du classement (augmentation, baisse ou stagnation) identifiée dans les rapports suivants permet de mesurer l'impact des actions entreprises et de juger de la réussite du projet de référencement. Cette première étape d'identification du besoin est

similaire à tout projet de conception de produits/services, néanmoins dans notre cas elle requiert obligatoirement la validation du référencieur. En effet, un projet de référencement ne débutera pas si l'ensemble des mots-clés n'a pas été validé au préalable par le référencieur.

			Position	Page
<input type="checkbox"/>	Mots-clés	Moteur	24 juin 2016	24 juin 2016
<input type="checkbox"/>	compensateurs de dilatation	 .fr (fr)	19	2
<input type="checkbox"/>	joint de dilatation	 .fr (fr)	59	6
<input type="checkbox"/>	manchette souple	 .fr (fr)	7	1
<input type="checkbox"/>	matelas d'isolation	 .fr (fr)	7	1
<input type="checkbox"/>	compensateur métallique	 .fr (fr)	5	1

Figure 14. Exemple d'un premier rapport de positionnement, "rapport point zéro", MyPoseo

3.2.2 Étape 2 : Étude de la faisabilité du projet

Après l'identification du besoin, le référencieur doit examiner le site web de son client afin d'identifier la technologie web employée (exemple : HTML5, Flash, etc.) et la plateforme utilisée (exemple : Wordpress, Joomla!, Prestashop, aucune plateforme, etc.). Si le site web du client est compatible avec l'implémentation des techniques de référencement, le référencieur peut commencer à proposer des actions correctives. Mais parfois le site web concerné ne permet pas l'implémentation des techniques de référencement. Par exemple, quand le site est développé en langage Flash [20] ou si la plateforme web ne permet pas la modification des balises HTML, ou si l'accès au code source est protégé en écriture. La meilleure solution dans ce cas de figure est de construire un nouveau site web en veillant à conserver le même nom de domaine pour éviter de perdre son ancienneté. Le référencieur doit donc être capable d'adapter ses méthodes de travail en fonction des contraintes qu'il aura identifiées.

3.2.3 Étape 3 : Proposition des actions

Quand le référencieur considère que le site web est compatible avec l'optimisation des critères du référencement, il doit ensuite proposer les meilleures actions pour améliorer le classement du site. Le choix des actions à implémenter dépend des données récoltées à partir de ses outils (exemple : audit HTML, analyse des liens retour, etc.) mais aussi grandement de l'expérience acquise par le référencieur avec ses

précédents projets. Ainsi, le référenceur proposera des actions qu'il a lui-même expérimentées sur d'autres clients comparables, ce qui rend cette tâche particulièrement artisanale³⁰. Ce travail ne concerne pas le client, il implique uniquement le référenceur. Le référenceur doit trouver un compromis entre le besoin du client, le respect des règles des moteurs de recherche et le degré d'acceptation du client concernant la modification de son site. Durant cette troisième étape, le client peut accepter ou non les actions proposées par le référenceur. Si le client refuse l'implémentation des actions, les résultats peuvent être décevants, le référenceur doit donc convaincre son client de l'importance et de l'efficacité de ses futures actions.

3.2.4 Étape 4 : Implémentation des actions

Si les actions proposées ont été acceptées par le client, le référenceur les implémentera sur le site web. Ce travail consiste à implémenter des critères *on-page* (exemple : insérer des mots-clés dans les balises HTML, créer des textes avec du contenu optimisé, créer de nouvelles pages, modifier la structure du site, etc.) ou des critères *off-page* (exemple : créer des liens retour de qualité depuis des sites web pertinents, créer un blog, créer des pages de réseaux sociaux, etc.). Quand toutes les actions ont été implémentées sur le site web, le référenceur réalise un dernier contrôle afin de vérifier qu'aucun problème ou aucune erreur ne perturbe l'amélioration du classement par la suite.

3.2.5 Étape 5 : Mesure de l'impact

Après l'implémentation des actions au sein du site web du client, le référenceur doit mesurer l'impact de ses actions en réalisant un suivi des résultats. L'évolution du positionnement doit être transmise au client pour qu'il puisse mesurer et juger l'efficacité du travail de référencement. Durant cette étape le référenceur envoie habituellement un rapport de positionnement au client à une fréquence d'environ un rapport par mois ce qui est une bonne

³⁰ Dans notre dernier chapitre, nous avons développé un système d'aide à la décision pour améliorer la rapidité et l'efficacité de cette étape.

fréquence pour percevoir une évolution du classement. Si les résultats s'améliorent, le contrat de référencement prend fin. Néanmoins si les résultats obtenus sont loin de ceux envisagés au départ, le référencier proposera de mettre en place de nouvelles actions. Parfois, le référencier proposera de modifier la liste initiale de mots-clés en ajoutant des mots-clés avec un meilleur potentiel. Cette dernière situation nécessite de recommencer le cycle du processus. Le processus de référencement est donc itératif jusqu'à la satisfaction du client.

Le diagramme de cas d'utilisation dans la figure 15 ci-dessous permet de représenter chacune des étapes du processus de référencement que nous avons décrites ainsi que les rôles des acteurs. Durant le processus de référencement, nous remarquons que le client participe uniquement à l'expression du besoin :

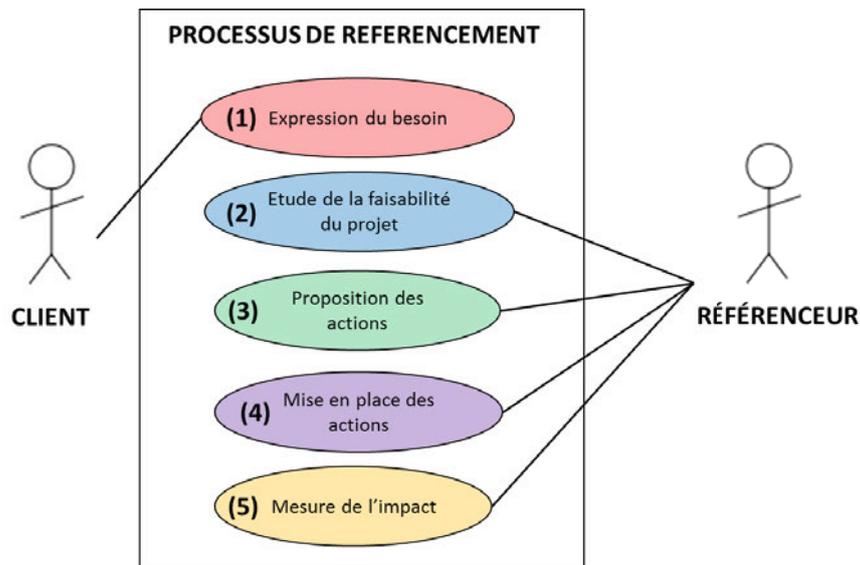


Figure 15. Diagramme des cas d'utilisation, processus de référencement

En décrivant en détail les cinq étapes du processus de référencement, nous avons pu mieux comprendre leur déroulement chronologique et identifier les parties prenantes. Par ailleurs, la similarité avec un processus de conception de produits/services est très forte puisque l'on retrouve quasiment les mêmes étapes : définition du problème, collecte des informations, génération de solutions, analyse et sélection des solutions, test et implémentation de la solution. Cette similarité permet de valider le recours à un méta-modèle issu de la conception de produits pour définir le processus de référencement.

Les acteurs humains sont amenés à interagir durant le processus de référencement. Il est nécessaire de bien comprendre ces interactions pour identifier leurs caractéristiques. Nous décrivons ces interactions dans la section suivante.

3.3 Interactions entre les acteurs du processus de référencement web

Dans la section précédente, nous avons décrit les cinq étapes du processus de référencement. Durant ces étapes le référenceur et le client interagissent pour mener à bien un projet de référencement. Après avoir étudié les échanges d'emails et les conversations téléphoniques entre les clients et les référenceurs au sein de notre entreprise pilote, nous décrivons ci-dessous comment interagissent habituellement ces acteurs durant le processus de référencement :

- (1) Identification du besoin :** pendant cette première étape, les acteurs humains collaborent peu. Néanmoins, quand un client souhaite conserver des mots-clés peu pertinents ou difficiles à référencer (exemple : concurrence élevée, thématique éloignée, etc.), le référenceur peut être amené à échanger fréquemment avec son client. Dans ce cas, l'objectif du référenceur est de convaincre le client de modifier son choix de mots-clés pour espérer obtenir de bons résultats. Dans de rares cas, si le client ne souhaite pas modifier sa liste de mots-clés (jugée non pertinente par le référenceur), un conflit [49] peut apparaître. En effet, un référenceur ne commencera pas un projet de référencement s'il juge que les mots-clés choisis ne permettront pas d'atteindre les objectifs fixés.

- (2) Étude de la faisabilité du projet :** pendant cette seconde étape, si le référenceur constate que le site web du client est compatible avec la mise en place des techniques de référencement, il passera à l'étape suivante, presque aucun échange n'aura lieu avec le client. Néanmoins si le référenceur constate que le site web du client n'est pas compatible avec la mise en place des actions de référencement, celui-ci devra convaincre le client de construire un nouveau site web. Si le client n'accepte pas, un conflit peut apparaître.

- (3) Proposition des actions :** durant cette troisième étape, le client peut accepter ou non l'implémentation des actions sur son site web. Si le client accepte, il y aura peu d'échanges entre les acteurs humains. Néanmoins si le client n'accepte pas, le résultat peut être décevant, ainsi le référenceur doit convaincre le client de l'importance et de l'efficacité de ses futures actions et justifier leur mise en place.
- (4) Implémentation des actions :** pendant cette quatrième étape, il y a très peu d'échanges avec le client excepté concernant l'échange des codes d'accès du site web pour réaliser l'implémentation des actions.
- (5) Mesure de l'impact :** durant cette dernière étape du processus de référencement, si les résultats sont bons, les échanges entre les acteurs s'arrêtent. Si les résultats ne sont pas à la hauteur, le référenceur doit reconsidérer l'ensemble des étapes, le processus recommence.

Cette étude nous a permis de constater qu'à l'exception de cas particuliers (manque de souplesse de la part du client), les interactions entre les acteurs sont limitées, peu fréquentes, et qu'elles font très peu intervenir le client dans le processus de référencement. La collaboration avec le client est donc plutôt faible lors d'un projet de référencement. Mais est-ce toujours vrai ? Qu'en est-il avec un client expérimenté ? Est-ce que le niveau de collaboration peut redéfinir notre schéma initial de processus de référencement ? Nous répondrons à ces questions dans la section suivante.

4. Collaboration et niveau de connaissance

Le méta-modèle présenté au début de ce chapitre nous a permis d'identifier une première version du processus de référencement. Cependant, lors de nos travaux de recherche, nous nous sommes aperçu qu'il n'existait non pas une, mais plusieurs versions de ce processus [37]. En effet, nous considérons que chaque version du processus de référencement se distingue par le niveau de collaboration entre les acteurs humains. L'activité de conception du référencement, tout

comme l'activité de conception de produits/services est à la fois **distribuée** [50] et **collective** [51]. Dès lors, elle nécessite la **coopération et la coordination de divers acteurs métier** porteurs de **connaissances propres** à leur domaine [52]. Par ailleurs, les acteurs d'un projet de référencement sont **cognitivement dépendants**. C'est-à-dire que le client ne peut réduire l'incertitude qui pèse sur la conduite de son activité (le référencement) sans recourir à un tiers (le référencier) [53]. En d'autres termes le client a besoin des connaissances et de l'expérience du référencier pour conduire l'activité de référencement. Du fait de leur différence de connaissance, les acteurs du projet doivent trouver un **référentiel commun** pour collaborer. Cela permet d'ajuster les décisions de chacun en fonction des connaissances de l'autre [54]. Le référencier doit ainsi ajuster le niveau de son discours en fonction des connaissances de son client afin que ce dernier puisse collaborer au mieux. Avec un client expérimenté, le référencier pourra conserver un niveau de langage élevé.

Cette approche nous a permis d'identifier deux niveaux de collaboration entre les acteurs d'un projet. Le premier niveau se traduit par une collaboration faible comme observée dans la section 3.3. Le deuxième niveau se traduit par une collaboration forte entre les acteurs, comme c'est le cas avec un client expérimenté. Ainsi, nous considérons que le niveau de connaissance du client influence le niveau de collaboration entre les acteurs. La littérature va dans ce sens et considère qu'un niveau de connaissance bas peut affaiblir voire supprimer l'aspect collaboratif du projet ce qui peut impacter son succès [5].

Nous allons voir que la réussite d'un projet de référencement peut ainsi pâtir d'une différence de connaissances trop importante entre les acteurs. En effet, durant un projet de référencement, le client est mis à contribution pour des étapes critiques présentées ci-dessous :

- **Le choix des mots-clés** : même appuyé par les conseils du référencier, le choix des mots-clés est une décision critique à prendre et elle détermine la réussite du projet. Un mauvais choix de mots-clés peut entraîner des délais très longs et des résultats médiocres. Ainsi, un client expérimenté, conscient de

l'importance du choix des mots-clés, pourra trouver des compromis et permettra au référenceur d'entamer un projet ayant de fortes chances d'aboutir sur des résultats satisfaisants.

- **La maîtrise du site web** : le niveau de maîtrise du site web par le client peut impacter la réussite du projet. Un client ayant une bonne connaissance technique de son site web sera à même d'ajouter des pages ou des contenus texte facilement, ce qui sera bénéfique pour la réussite du projet. À l'inverse un client n'ayant aucune maîtrise de son site obligera le référenceur à se charger lui-même des intégrations. Dès lors, le référenceur devra découvrir le fonctionnement du site web et en fonction de la complexité du site, cela pourra allonger les délais du projet. Par ailleurs, certains clients n'ont même pas la main sur leur site web ce qui peut bloquer l'implémentation des actions et reconsidérer l'issue du projet.
- **Le degré d'acceptation des modifications** : le client est propriétaire du site web, dès lors il est en capacité d'accepter ou non certaines modifications. Un client inexpérimenté aura tendance à ne pas vouloir changer l'aspect de son site (exemple : ajout de texte, modification des titres, etc.) et privilégiera le visuel. À l'inverse, un client expérimenté, conscient des enjeux techniques, trouvera un compromis pour préserver l'efficacité des actions de référencement et l'identité visuelle du site.

4.1 Deuxième version du processus de référencement web

Nous avons vu que le client peut être amené à contribuer au déroulement d'un projet de référencement et que son niveau de connaissance pouvait impacter la réussite de ce projet. Un client expérimenté sera plus à même de trouver des compromis ce qui permettra d'améliorer les chances de succès et les délais du projet contrairement à un client inexpérimenté. Ces observations nous amènent à reconsidérer notre première version du processus de référencement.

La figure 16 qui suit présente une deuxième version du processus de référencement dans lequel le client est expérimenté. Il possède un niveau de connaissance élevé de l'activité de référencement.

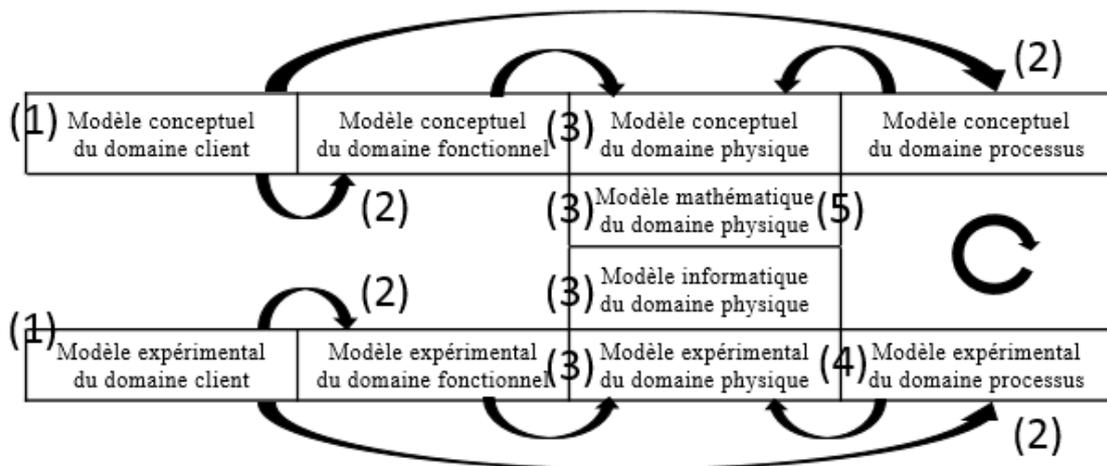


Figure 16. Deuxième version du processus de référencement, cas d'un client expérimenté

Nous constatons que le processus de référencement présenté ci-dessus diffère de la première version, car il intègre désormais le modèle expérimental pour les étapes **(1) Identification du besoin** et **(2) Étude de la faisabilité du projet**. En effet, le client expérimenté a déjà réalisé des tests et des essais de son côté ce qui lui a permis d'acquérir une première expérience de l'activité de référencement.

Nous pouvons aussi constater que le client expérimenté intervient désormais pour chaque étape du processus, dont voici un descriptif :

Étape 1, identification du besoin : le client expérimenté a connaissance des exigences concernant le choix des mots-clés (ex : concurrence, thématique, etc.), il sera donc plus à même de collaborer avec le référenceur pour trouver un compromis et définir une liste de mots-clés pertinente.

Exemple : Pour une activité de visite de Paris en 2CV, le client expérimenté proposera de lui-même les mots-clés « visiter paris », « visiter paris en 2CV », « visiter paris 2CV », et évitera les mots-clés trop génériques/concurrentiels tels que « paris » ou « 2CV ».

Étape 2, étude de la faisabilité du projet : le client expérimenté a une bonne connaissance de son site web et il sait comment l'administrer. Si le site permet la mise en place de techniques de référencement, le référenceur en sera rapidement informé par le

client. Les délais de l'étude de la faisabilité du projet seront donc considérablement raccourcis.

Exemple : le référenceur demande au client expérimenté si les balises <title> peuvent être modifiées sur son site. Le client possède le CMS Prestashop et il sait que les balises <title> sont modifiables sur toutes les pages du site. Le projet de référencement est donc faisable.

Étape 3, proposition des actions : le client expérimenté donne directement son aval au référenceur concernant les solutions proposées et les actions à implémenter. Des compromis peuvent être trouvés facilement concernant les modifications à réaliser, car le client a une bonne maîtrise technique de son site.

Exemple : le référenceur propose au client expérimenté d'ajouter du contenu texte incluant des mots-clés importants sur sa page d'accueil. Le client expérimenté valide cette proposition et confirme qu'il pourra ajouter un paragraphe de texte sur sa page d'accueil ce qui permet de valider la proposition du référenceur.

Étape 4, mise en place des actions : le client expérimenté implémente lui-même les balises et les ajouts de textes sur son site web. Il sait comment procéder et cela fait gagner du temps au référenceur, qui n'aura plus qu'à valider ce travail.

Exemple : le client expérimenté intègre lui-même les balises <h1> rédigées par le référenceur au sein de son site web. Une fois que tout a été intégré, le référenceur vérifie que les balises sont correctement intégrées (pas de balise <h1> en double, pas de balise <h1> manquante, etc).

Étape 5, mesure de l'impact : le client sait interpréter les résultats issus des rapports de positionnement. En accord avec les décisions du référenceur, il peut même corriger certains aspects bloquant le gain de position.

Exemple : le client expérimenté s'aperçoit que la position d'un mot-clé stagne depuis plusieurs mois. Il constate que la page dédiée à ce mot-clé comporte une seule occurrence du mot-clé, il décide donc d'en rajouter dans le texte.

Le diagramme de cas d'utilisation sur la figure 17 ci-dessous permet d'identifier chacune des étapes de la deuxième version du processus

de référencement que nous venons de décrire ainsi que les rôles des acteurs. Nous pouvons remarquer que le client expérimenté collabore désormais avec le référencieur pendant toutes les étapes du projet.

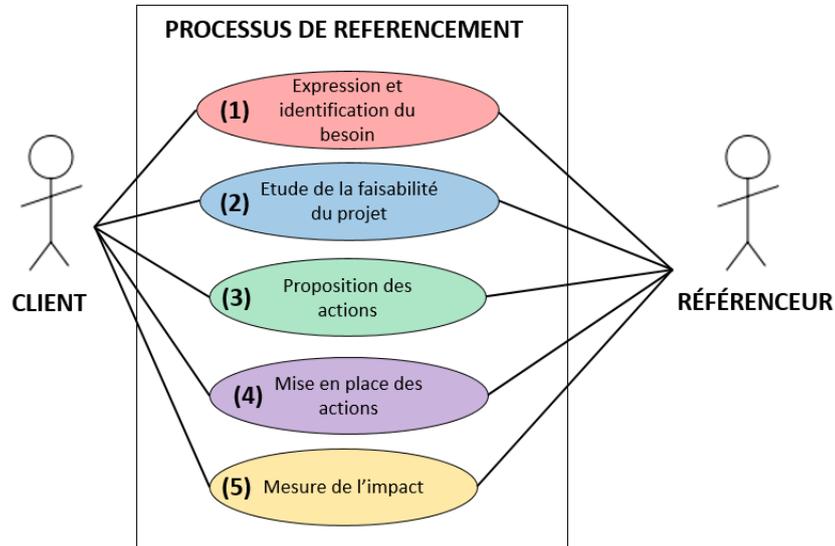


Figure 17. Diagramme des cas d'utilisation, seconde version du processus de référencement collaboratif

La collaboration avec un client expérimenté permet au référencieur de réduire les délais et d'améliorer la réussite d'un projet. Cependant il faut être conscient que tous les clients ne sont pas expérimentés. Ainsi, le processus de référencement doit être adapté de façon à rester efficace pour tous les types de clients.

5. Discussion et conclusion

La mise en oeuvre d'une démarche de conception de produits pour concevoir le référencement nous a permis d'avoir recours à un méta-modèle d'ingénierie. Grâce à ce meta-modèle issu de la conception de produits nous avons pu définir une première version du processus de référencement. En identifiant les acteurs, leurs interactions et les cinq étapes qui constituent ce processus, nous sommes parvenus à le structurer. Nous avons ainsi pu répondre à notre première hypothèse :

Hypothèse n°1 : « *Le processus de référencement est un processus complexe et structurable* »

La modélisation est le langage de la conception. Le référencement est une activité de conception. Ainsi, pour maîtriser sa complexité, nous avons souhaité concevoir le référencement comme un « processus ». Au départ ce processus n'existait pas, car le référencement n'était pas considéré comme une activité d'ingénierie. Le méta-modèle d'ingénierie utilisé nous a permis de mettre en évidence des modèles et des domaines pour structurer cette activité, et ainsi définir les étapes clés du processus de référencement. En analysant les étapes du processus de référencement, nous avons également pu mettre en évidence le fait que les interactions entre les acteurs sont peu fréquentes ce qui induit une collaboration presque inexistante. Néanmoins, dès lors que le client est expérimenté, la collaboration s'intensifie ce qui a pour effet de modifier le déroulement du processus de référencement, ses délais voire sa réussite. En observant l'activité de référencement, nous sommes parvenus à définir une seconde version du processus de référencement portant à deux le nombre de schémas possibles. Lors de nos recherches, nous avons fait le choix de prendre en compte l'aspect *client*, et notamment son *niveau de connaissance* pour concevoir le processus de référencement. Si le rôle du client est rarement évoqué dans la littérature [37], nous considérons de notre côté que cette approche est nécessaire. En effet, que le client soit ou non expérimenté, son rôle ne se limite pas à celui d'observateur. Il doit être un véritable acteur du projet et il doit collaborer avec le référencier concernant certaines décisions touchant les aspects décisifs du projet, comme par exemple le choix des mots-clés ou le degré de modification du site. Ainsi le développement d'une **démarche de travail participative** [5] permettant aux acteurs de collaborer efficacement, quelque soit le niveau de connaissance du client [55] permettrait aux acteurs du projet de devenir **cognitivement inter-dépendant**, c'est-à-dire qu'ils auraient besoin l'un de l'autre pour conduire l'activité de référencement [53].

Le développement d'un **outil de support pour la prise de décision** et servant de **référentiel commun** entre les acteurs du processus de référencement est une des solutions à notre problème. Dans le prochain chapitre, nous modélisons le processus de référencement à partir de systèmes multi-agents, pour approfondir cet axe de recherche.

Chapitre 4. Modélisation du processus de référencement web

1.	<i>Introduction</i>	70
2.	<i>Le paradigme agent pour modéliser le processus de référencement web</i>	71
2.1	Le référencement web : un système distribué	71
2.2	Choix de l’outil de modélisation : les systèmes multi-agents	72
3.	<i>Un modèle d’agents du processus de référencement web</i>	74
3.1	Le modèle	74
3.2	Modèle des agents SEO	76
3.3	Propriété des agents SEO	76
3.4	Comportement d’un agent SEO	77
3.5	Coopérations et interactions	78
4.	<i>Modélisation agents du processus de référencement web</i>	80
4.1	Organisation des agents	82
4.1.1	Les agents besoin	82
4.1.2	Les agents fonction	82
4.1.3	Les agents contrainte	82
4.1.4	Les agents solution	82
4.2	Des agents solution : Page et Critère	83
5.	<i>Discussion et conclusion</i>	86

1. Introduction

Dans le chapitre précédent, nous avons pu mettre en évidence que le processus de référencement pouvait être décliné en plusieurs versions grâce à l’utilisation d’un méta-modèle d’ingénierie, ce qui nous a permis d’en décliner deux versions. La décomposition du processus en plusieurs étapes nous aussi a permis d’identifier les acteurs, leurs interactions ainsi que leurs niveaux de collaboration. C’est ce dernier facteur qui permet de faire la distinction entre les deux versions de notre processus. Par ailleurs, nous avons remarqué que le niveau de collaboration entre les acteurs est influencé par le niveau de connaissance du client. Avec un client expérimenté, le référenceur aura plus de facilité à collaborer. À l’inverse un niveau de collaboration faible peut rallonger les délais, voire nuire à la réussite d’un projet de référencement. Nous avons conclu que pour permettre une collaboration efficace avec le référenceur et ainsi obtenir des résultats satisfaisants, il est nécessaire d’adopter une démarche collaborative [5].

Après avoir défini le processus de référencement, nous cherchons désormais à le modéliser. L'objectif de ce chapitre est de répondre à notre deuxième hypothèse :

Hypothèse n°2 : « *Le processus de référencement est modélisable* »

Sur la base des premières modélisations du chapitre 3, nous nous proposons ici de modéliser plus formellement le processus de référencement. En effet, nous avons identifié des acteurs collaborant à différents niveaux, des étapes clés du processus ainsi que des interactions. Le paradigme agent peut être utilisé pour modéliser plus finement et plus formellement ce processus. Par ailleurs, cette modélisation agent nous permettra dans le prochain chapitre de concevoir un système d'aide à la décision servant de support à l'activité de référencement.

Ce chapitre est composé de quatre sections. La première section explique pourquoi les systèmes multi-agents sont adaptés pour modéliser le processus de référencement. Dans la deuxième section, nous présentons notre modèle d'agent, les caractéristiques et les propriétés des agents impliqués dans ce modèle. La troisième section expose la modélisation de notre modèle d'agents. Nous terminons ce chapitre par une discussion et une conclusion.

2. Le paradigme agent pour modéliser le processus de référencement web

2.1 Le référencement web : un système distribué

Comme nous l'avons exposé dans le chapitre 3, le processus de référencement implique la coopération entre plusieurs entités autonomes qui interagissent ensemble pour concevoir le référencement. Ces entités sont au nombre de quatre : les *besoins*, les *fonctions*, les *solutions* et les *contraintes*. Le processus de référencement est également composé d'acteurs, le client et le référencier qui interagissent, voire collaborent, pour aboutir à la réussite du projet. Les interactions au sein du processus de référencement peuvent donc être de trois types différents. Chacune de ces interactions est illustrée sur la figure 18 ci-dessous :

(1) Entre les **acteurs du projet** (ex : client et référencier)

- (2) Entre **acteurs et entités** (ex : référencieur et *contraintes*)
- (3) Entre les **entités** (ex : *fonctions* et *solutions*)

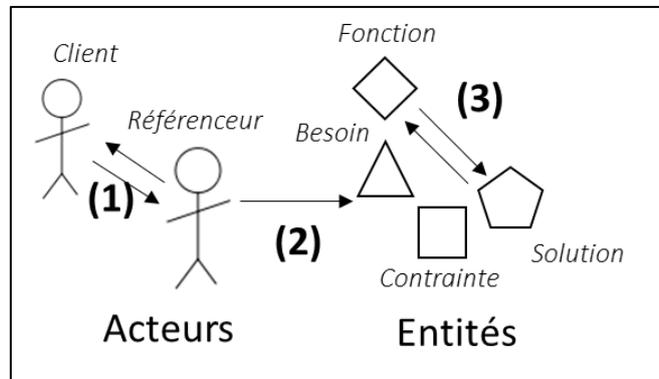


Figure 18. Interactions entre acteurs et entités du système

Comme un processus de conception classique, le processus de référencement est constitué d'acteurs et d'entités qui « interagissent » afin d'élaborer une solution adaptée. Ainsi, pour modéliser le processus de référencement, nous le considérons comme un système « *distribué* ». En effet, un système distribué contient un ensemble d'entités autonomes de calcul interconnectées qui peuvent communiquer entre elles [56]. Cette considération nous semble adaptée, car nous savons que la conception est une activité distribuée [51] ; par ailleurs cela permet d'isoler les entités afin de mettre en évidence leurs interactions. Dans le chapitre 2 nous avons constaté qu'en raison de sa structure dynamique et évolutive, de la connaissance limitée des méthodes et des critères de classement, du volume et de la diversité des données, du nombre important d'interactions et de l'incertitude des résultats, le référencement est un problème complexe. Les systèmes distribués sont tout à fait adaptés pour résoudre des problèmes complexes [56].

2.2 Choix de l'outil de modélisation : les systèmes multi-agents

Le paradigme agent permet par essence la modélisation de systèmes distribués [57], [58]. En effet, un système multi-agents (SMA) est un système informatisé composé d'agents intelligents qui interagissent dans un environnement [56], [59]–[62]. L'approche agent est très efficace pour la résolution de problèmes complexes [61], [62] et plus généralement pour assister et aider à la conception de produits/services [63]. En effet, les propriétés des agents (autonomie,

distribution, adaptabilité, flexibilité et coopération) [56], [61], [62], [64] permettent de faciliter les échanges d'informations entre des entités.

D'autres outils de modélisation [65], [66] existent comme les réseaux de neurones artificiels, les algorithmes génétiques ou les réseaux de Petri. Après une comparaison que nous illustrons dans le tableau 2, il s'est avéré que le paradigme agents était le mieux adapté pour répondre à nos attentes en termes de modélisation [67]. Ce qui n'est guère surprenant pour un paradigme apparu dans le domaine de l'intelligence artificielle distribuée [56].

Tableau 2. Comparatif sur l'adéquation d'outils courants de modélisation de systèmes aux caractéristiques du processus de référencement

Caractéristiques du processus de référencement	Système multi-agents	Réseau de neurones	Algorithme génétique	Réseau de Petri
Distribution	importante	importante	moyenne	importante
Complexité	importante	importante	importante	moyenne
Autonomie	importante	moyenne	faible	moyenne
Flexibilité/Adaptabilité	importante	moyenne	importante	faible
Coopération	importante	faible	pas adaptés	faible

Nous détaillons ci-dessous l'adéquation des systèmes multi-agents aux principales caractéristiques du processus de référencement présentées dans le précédent tableau :

- **Distribution** : dans un système multi-agents chaque agent est une entité spécifique interagissant avec les autres agents ce qui permet de décomposer un système ou un processus, et d'en proposer une architecture distribuée [61], [62], [64]
- **Complexité** : les problèmes complexes ont la particularité d'associer des éléments aux interactions multiples, et d'avoir un environnement non nécessairement restreint ; les systèmes multi-agents contrairement aux systèmes classiques peuvent évoluer dynamiquement pour résoudre ce type de problèmes [68].
- **Autonomie** : un agent est une entité qui possède ses propres objectifs, il peut spontanément prendre des initiatives pour les atteindre [69]

- **Flexibilité/Adaptabilité** : un agent est capable de contrôler et d'adapter ses aptitudes (décisionnelles, comportementales, etc.) en réponse aux connaissances internes ou aux changements de l'environnement [56].
- **Coopération** : les agents sont capables de coopérer entre eux pour trouver une solution adéquate et réaliser une tâche ou un objectif commun [61], [62], [64], [70].

Ces caractéristiques nous permettent de proposer l'utilisation d'agents pour modéliser le processus de référencement. Par ailleurs, les recherches sur les systèmes multi-agents sont importantes dans notre laboratoire [45], [48], [71], [72] ce qui nous conforte dans le choix de cet outil.

Dans cette section nous avons montré que les systèmes multi-agents de par leurs caractéristiques (distribution, autonomie, flexibilité, coopération) permettent la résolution de **problèmes complexes, distribués** et plus spécifiquement de **problèmes de référencement**. D'autre part leurs caractéristiques permettent d'aboutir à des raisonnements, mais aussi à des prises de décisions. C'est dans cette optique que nous avons élaboré notre modèle d'agents présenté dans la section suivante.

3. Un modèle d'agents du processus de référencement web

3.1 Le modèle

Fougères [73] a proposé un modèle de systèmes multi-agents qui a été utilisé pour des modélisations en conception de produits [48] et que nous avons choisi de réutiliser pour modéliser le processus de référencement. Pour rester en adéquation avec notre processus, notre système multi-agents est composé de quatre communautés d'agents : *la communauté des agents besoin, la communauté des agents fonction, la communauté des agents contrainte et la communauté des agents solution.*

Des interactions entre les communautés d'agents, les agents eux-mêmes et les acteurs humains se déroulent durant le processus de référencement. Ainsi, les *agents besoin* réceptionnent et identifient les

besoins exprimés par le client. Les *agents contrainte* reçoivent et identifient les contraintes identifiées par le référencieur. Les *agents fonction* réceptionnent les messages des agents *besoin* pour les transformer en fonctions. Les *agents solution* reçoivent les messages des *agents fonction et contrainte* pour ensuite générer des solutions pertinentes.

Notre modèle d'agents est illustré sur la figure 19 ci-dessous. Les flèches représentent les interactions déjà identifiées sur la figure 18 de ce chapitre.

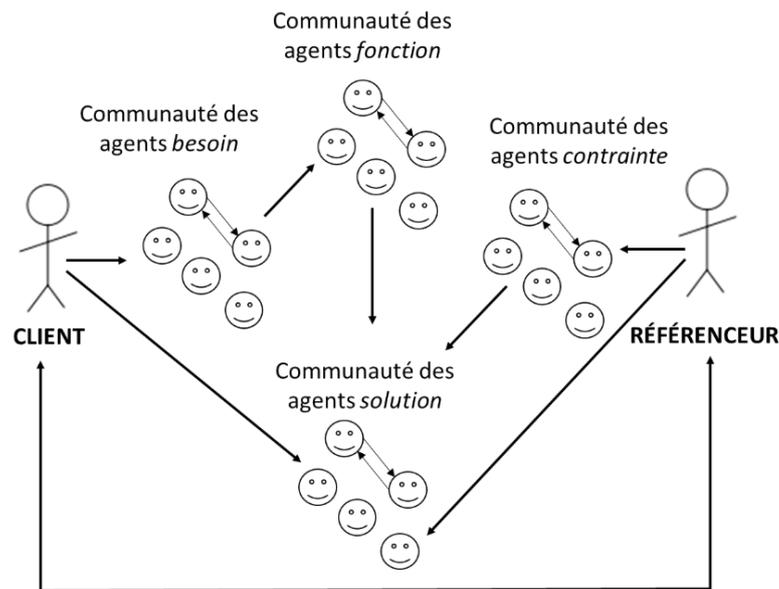


Figure 19. Modèle d'agents du processus de référencement

Le système multi-agents que nous avons développé pour modéliser le processus de référencement est décrit par le quadruplet suivant (1) :

$$M = \langle A, I, P, O \rangle \quad (1)$$

où A représente l'ensemble des *agents* SEO, I représente l'ensemble des *interactions* entre les agents, P est l'ensemble des *rôles* joués par les agents SEO, et O est l'ensemble des *organisations* d'agents au sein des communautés. En effet, en accord avec le méta-modèle d'ingénierie décrit au chapitre 3, chaque agent appartient à l'une de ces quatre communautés : $\langle Besoin \rangle$, $\langle Fonction \rangle$, $\langle Contrainte \rangle$ et $\langle Solution \rangle$. Mais ils peuvent aussi appartenir à des sous-communautés, comme nous le verrons dans la section 4.2.

3.2 Modèle des agents SEO

Les agents SEO sont au cœur de notre modèle et possèdent plusieurs fonctions. Nous considérons qu'un agent SEO $\alpha_i \in A$ peut se décrire selon le quadruplet suivant (2) :

$$\alpha_i = \langle \Phi_{\Pi(\alpha_i)}, \Phi_{\Delta(\alpha_i)}, \Phi_{\Gamma(\alpha_i)}, K_{\alpha_i} \rangle \quad (2)$$

ou $\Phi_{\Pi(\alpha_i)}$ est la fonction *d'observation* d'un agent SEO α_i ; $\Phi_{\Delta(\alpha_i)}$ est la fonction de *décision* d'un agent SEO α_i ; $\Phi_{\Gamma(\alpha_i)}$ est la fonction *d'action* d'un agent SEO α_i ; K_{α_i} est l'ensemble fini des *connaissances* de l'agent α_i - la connaissance des agents intègre les valeurs du domaine, les connaissances directes (le réseau d'affinité, particulièrement à l'intérieur des communautés d'agents), les connaissances dynamiques (événements observés, état interne), et les règles de décision.

Nous décrirons et illustrerons davantage les règles de décision au chapitre suivant. Nous montrerons qu'elles respectent le formalisme ECA : [SI évènement(s) avec condition(s) ALORS action(s)], mais que certaines utilisent la logique floue.

3.3 Propriété des agents SEO

Tout comme dans le modèle original proposé par [73], les agents SEO possèdent trois propriétés distinctes :

- **Propriété 1 (perception)** : les agents ont leur propre niveau de connaissance, ils peuvent partager et communiquer leur connaissance. Ils sont aussi capables de percevoir des informations produites ou communiquées par des acteurs humains ou d'autres agents.
- **Propriété 2 (interaction)** : les agents ont un degré d'affinité avec les autres agents, ils peuvent approuver ou désapprouver les messages reçus et ils sont capables d'interagir avec leur environnement de manière intra-communautaire ou inter-communautaire

(avec des agents de leurs communautés ou avec des agents d'autres communautés).

- **Propriété 3 (adaptation)** : les agents sont capables de s'adapter à l'évolution de l'environnement, ils sont capables de modifier leurs interactions avec les autres agents et d'ajuster leur comportement en fonction de règles de décision auxquelles ils obéissent.

3.4 Comportement d'un agent SEO

Les agents SEO peuvent interagir entre eux ou avec des acteurs humains. Ainsi, en situation de fonctionnement, un agent SEO réalisera trois étapes entre le moment où il reçoit une information et le moment où il agira. L'exemple de scénario ci-dessous permet d'identifier ces trois étapes :

- (1) L'agent SEO α_i **reçoit un message** provenant d'un autre agent ou d'un acteur humain (ex : « *référencer le site web sur le mot-clé KW_i* »).
- (2) Ensuite l'agent SEO α_i , en fonction de ses connaissances, va **prendre une décision** (ex : « *ajouter le nouveau mot-clé KW_i* »).
- (3) Puis l'agent SEO α_i va **agir** en proposant à un agent ou un acteur humain d'ajouter le mot-clé KW_i .

La figure 20 qui suit présente l'architecture fonctionnelle d'un agent SEO α_i dont nous venons d'identifier les trois étapes de fonctionnement :

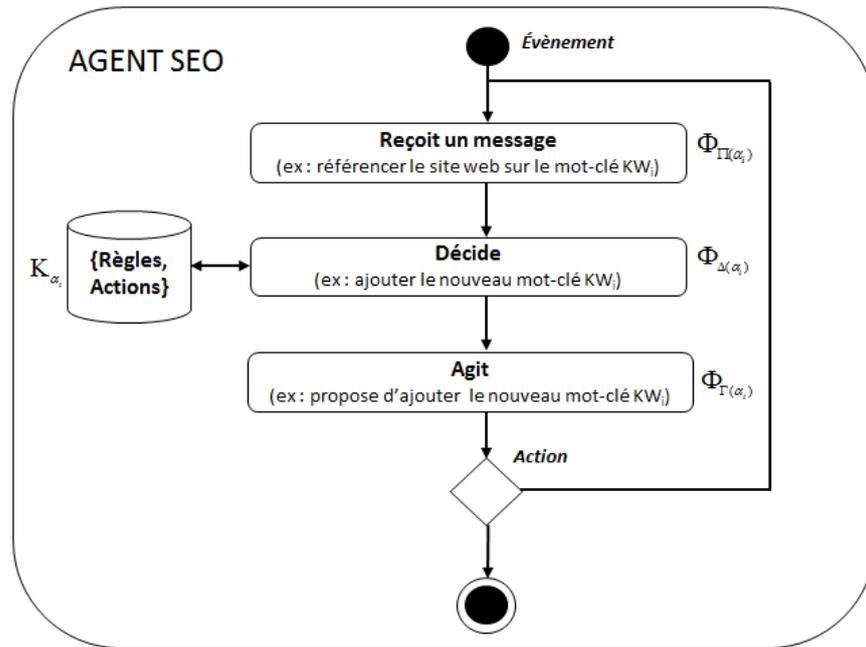


Figure 20. Architecture fonctionnelle d'un agent SEO

3.5 Coopérations et interactions

Dans notre modèle, un agent est capable d'interagir avec son environnement. Une interaction $l_i \in I$ entre deux agents SEO α_s and α_d est définie par le triplet suivant (3) :

$$l_i = \langle \alpha_s, \alpha_d, \gamma_c \rangle \quad (3)$$

ou $\alpha_s \in A$ et $\alpha_d \in A$ sont les agents SEO source et destinataire de l'interaction, et $\gamma_c \in \Gamma$ est un acte de coopération. La figure 21 représente un échange entre deux agents SEO.

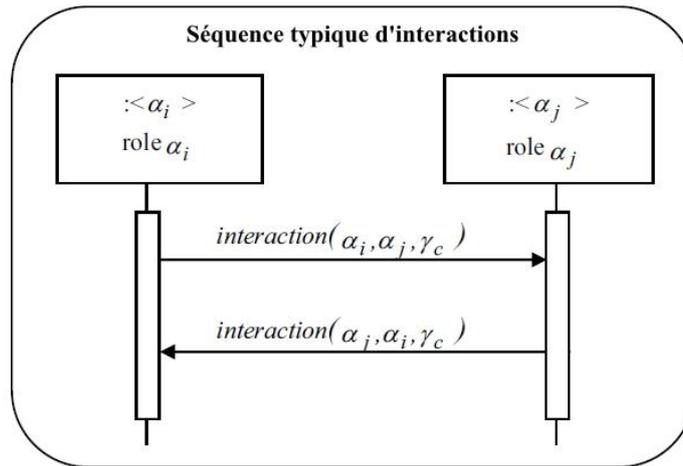


Figure 21. Représentation d'un échange entre deux agents SEO (double interaction)

Sur cette figure l'agent α_i interagit avec l'agent α_j selon un acte de coopération γ_c . Dans notre modèle, les agents SEO sont capables de réaliser neuf actes de coopération (4) dont nous donnons la description dans le tableau 3 ci-dessous.

$$\Gamma = \left\{ \begin{array}{l} \text{inform, reply, diffuse, propose, init} \\ \text{ask, accept, refuse, acknowledge} \end{array} \right\} \quad (4)$$

Tableau 3. Table descriptive des actes de coopérations des agents SEO

Acte de coopération	Descriptif	Syntaxe
<i>inform</i>	Un agent est capable d'informer un autre agent à propos de sa valeur, par exemple son âge	<i>inform(agent1, agent2, agent1value=x)</i>
<i>reply</i>	Un agent peut répondre à un autre agent à propos de sa valeur, par exemple son âge	<i>reply(agent2, agent1, agent2value=y)</i>
<i>diffuse</i>	Un agent peut diffuser une information à plusieurs agents	<i>diffuse(agent1, [listeagents], agent1value=x)</i>
<i>propose</i>	Un agent/humain peut proposer une information à un autre agent/humain	<i>propose(agent1, agent2, agent1value=z)</i>
<i>init</i>	Un agent/humain peut initialiser la valeur d'un agent	<i>init(agent1, x)</i>
<i>ask</i>	Un humain/agent peut demander à un humain/agent de réaliser une action	<i>ask(agent1, agent2, actioni)</i>
<i>accept</i>	Un humain/agent peut valider la réalisation d'une action à un humain/agent	<i>accept(agent1, agent, actioni)</i>
<i>refuse</i>	Un humain/agent peut refuser la réalisation d'une action à un humain/agent	<i>refuse(agent1, agent, actioni)</i>
<i>acknowledge</i>	Un humain/agent peut confirmer la réalisation d'une action à un humain/agent	<i>acknowledge(agent1, agent2, actioni)</i>

Comme nous l'avons présenté, plusieurs actes de coopération sont possibles, que ce soit entre agents, entre acteurs humains ou entre agents et acteurs humains. Des interactions sont aussi possibles entre des agents de communautés différentes. Au total, nous avons identifié trois types d'interactions :

- Entre les acteurs humains (les acteurs du projet)
- Entre les agents de la même communauté (interactions intra-communautaires)
- Entre les agents provenant d'autres communautés (interactions inter-communautaires)

Dans nos expérimentations nous n'avons pas travaillé sur les interactions entre les acteurs humains, celles-ci ne sont donc pas modélisées dans notre modèle.

4. Modélisation agents du processus de référencement web

Dans le chapitre 3 nous avons montré que le niveau de collaboration entre le client et le référencier pouvait influencer la réussite d'un projet de référencement. Nous avons déduit que pour améliorer la réussite et les délais d'un projet de référencement, nous devons privilégier une approche faisant participer le client durant toutes les étapes du projet. Le processus de référencement à modéliser doit aussi être pertinent, c'est à dire qu'il doit proposer des solutions adéquates en fonction des besoins du client et des contraintes identifiés par le référencier. Cette approche est illustrée sur la figure 22 ci-dessous :

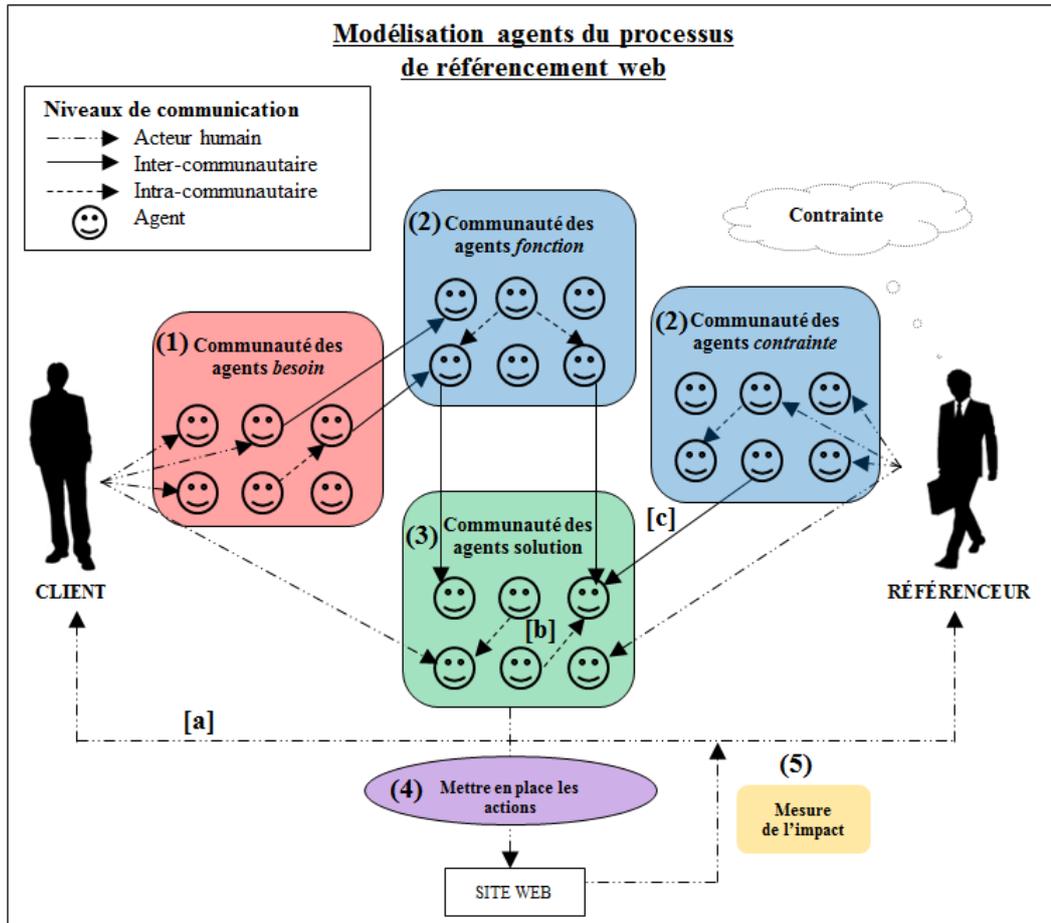


Figure 22. Modélisation agents du processus de référencement web

Dans cette modélisation agents, les entités sont représentées par des communautés d'agents : *<Besoin>*, *<Fonction>*, *<Contrainte>* et *<Solution>*.

Par ailleurs, notre modélisation reprend les cinq étapes du processus de référencement identifiées dans le chapitre 3, à savoir :

1. Identification du besoin (rouge)
2. Étude de la faisabilité du projet (bleu)
3. Proposition des actions (vert)
4. Implémentation des actions (violet)
5. Mesure de l'impact (jaune)

Enfin les lettres [a], [b] et [c] représentent respectivement les interactions entre acteurs humain, intra-communautaire, et intercommunautaire.

4.1 Organisation des agents

Comme nous l'avons évoqué ci-dessus, les agents issus de notre modèle appartiennent tous à l'une des quatre communautés suivantes : *besoins*, *fonctions*, *contraintes* et *solutions*. Nous décrivons les agents de chacune de ces communautés ci-dessous.

4.1.1 *Les agents besoin*

Les agents *besoin* réceptionnent et identifient les besoins exprimés par le client. Ce sont les seuls agents en contact avec le client. Le besoin d'un client est constitué de deux éléments : le/les mot(s)-clé(s) à référencer et la position désirée, par exemple « *dans les trois premières positions* » ou « *sur la première page* ». Une fois que les besoins du client sont réceptionnés, les agents *besoin* informent les agents *fonction*.

4.1.2 *Les agents fonction*

Les agents *fonction* réceptionnent les messages des agents *besoins* pour les transformer en fonctions. Leur intérêt est de formaliser la demande du client pour qu'elle puisse être compréhensible par le système. Une fois que les besoins sont transformés en fonction, les agents *fonction* informent les agents *solution*.

4.1.3 *Les agents contrainte*

Les agents *contrainte* reçoivent les contraintes identifiées par le référencier. Le référencier peut identifier deux types de contraintes : un manque ou un problème. Un manque peut s'exprimer comme ceci : « *le mot-clé KWi n'est pas présent sur le site web* ». Un problème peut s'exprimer comme ceci : « *le contenu de la page Pi est dupliqué* ». Une fois que les contraintes sont réceptionnées, les agents *contrainte* informent les agents *solution*.

4.1.4 *Les agents solution*

Les agents *solution* reçoivent les messages des agents *fonction* et *contrainte*. Ensuite ils génèrent des solutions pertinentes pour répondre aux besoins du client tout en respectant les contraintes identifiées par le référencier. Dans nos travaux de recherche, nous nous sommes focalisés sur la communauté des agents *solution*, car c'est à ce niveau que se situent les prises de décisions. En effet, nous pensons que les agents *solution* peuvent servir de support à l'activité de référence-

ment. Nous développerons davantage cet axe de recherche dans la section suivante.

4.2 Des agents solution : Page et Critère

Nous considérons la page web comme un agent *solution* car la majorité des solutions peuvent être implémentées dans des pages web lors d'un projet de référencement. Une page web est composée de plusieurs critères. Pour améliorer le classement d'une page, le référencieur doit améliorer certains de ces critères. Comme nous l'avons vu, ils sont divisés en deux groupes, les critères *off-page* (exemple : nombre de backlinks) et les critères *on-page* (exemple : répétition d'un mot-clé). Comme les pages web, les critères sont également considérés comme des agents *solution*. La figure 23 ci-dessous représente les interactions entre les agents *solution* *Page*, *Mot-clé* et *Backlink*. Nous pouvons remarquer que les groupements d'agents *Page*, *Mot-clé* et *Backlink* forment des sous-communautés d'agent *solution*.

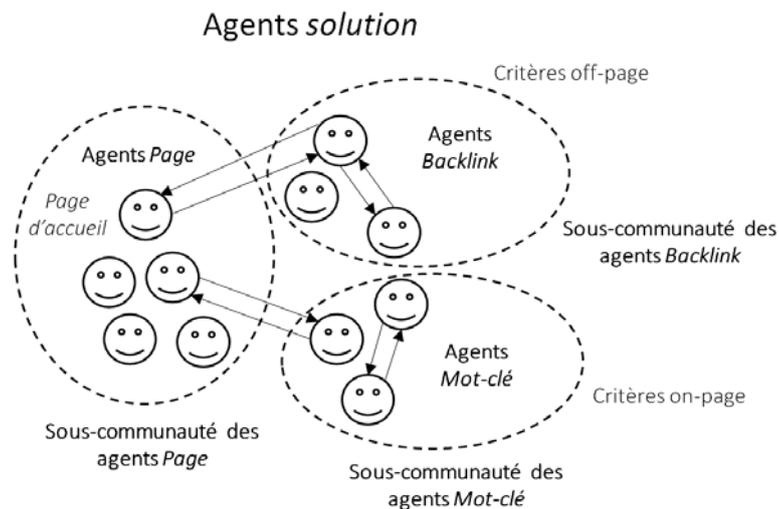


Figure 23. Agents solution : *Page*, *Mot-clé* et *Backlink*

Les agents *solution* peuvent avoir cinq types de connaissances dont nous donnons quelques exemples :

- **Connaissance de l'environnement** : un agent *Mot-clé* sait avec quels agents *Page* il interagit, il sait donc dans quelle page il apparaît.
- **Connaissance de soi-même** : un agent *Mot-clé* connaît son nombre de répétitions dans une page donnée.

- **Connaissance de ses propres compétences** : un agent *Page* sait qu'il peut demander à augmenter le nombre de répétitions d'un mot-clé, ainsi il peut déclencher une règle de décision pour informer le référenceur ou l'agent *Mot-clé* qu'il doit réaliser cette action.
- **Connaissance de contrôle de l'activité des agents** : un agent *Page* cherche à être informé fréquemment de son classement afin de savoir s'il doit déclencher (ou non) une règle de décision pour tenter de l'améliorer. Les règles de décision seront décrites au chapitre 5.
- **Connaissance des interactions avec les autres agents** : chaque agent sait à quelle communauté/sous-communauté il appartient et avec quelle communauté/sous-communauté il interagit. Un agent *Page* appartient à la sous-communauté des agents *Page*, il peut interagir avec un agent *Mot-clé* appartenant à la sous-communauté des agents *Mot-clé*.

Les agents *solution* sont aussi capables d'échanger des informations entre eux. Un agent *Page* peut donc interagir avec un agent *Mot-clé* et inversement. Durant le processus de référencement, ces interactions sont réalisées grâce à des échanges de messages. Prenons comme scénario un projet de référencement pour identifier ces messages :

- (1) Le client **demande** au référenceur d'améliorer son classement (exemple : « *J'aimerais que mon site web soit présent dans les trois premières positions dans les résultats de Google.fr sur le mot-clé KW_i* »)
- (2) Le référenceur fait deux observations : (a) le mot-clé KW_i n'est pas présent sur le site web, (b) le mot-clé KW_i ne peut pas être inséré sur les pages actuelles du site, car sa thématique est trop éloignée de la thématique générale du site. Pour respecter les règles du moteur de recherche, le référenceur prend la décision de créer une nouvelle page pour isoler le mot-clé KW_i . Il **initialise** la valeur de l'agent *besoin* (exemple : « ajouter le nouveau mot-clé KW_i ») et la valeur de l'agent *contrainte* (exemple : « ajouter une nouvelle page P_i »).

- (3) Après cela, l'agent *besoin informe* l'agent *fonction* de la création du nouveau mot-clé KW_i
- (4) Puis l'agent *contrainte informe* l'agent *solution* de créer une nouvelle page P_i . Ensuite l'agent *fonction informe* l'agent *solution* d'insérer le mot-clé KW_i sur la page P_i .
- (5) Enfin les agents *solution informent* le référencier de la création de la page P_i qui contient le mot-clé KW_i . Finalement, le référencier *propose* cette action au client qui peut *accepter* (ou non) de l'implémenter.

Le diagramme de séquence présenté dans la figure 24 illustre cet échange de messages :

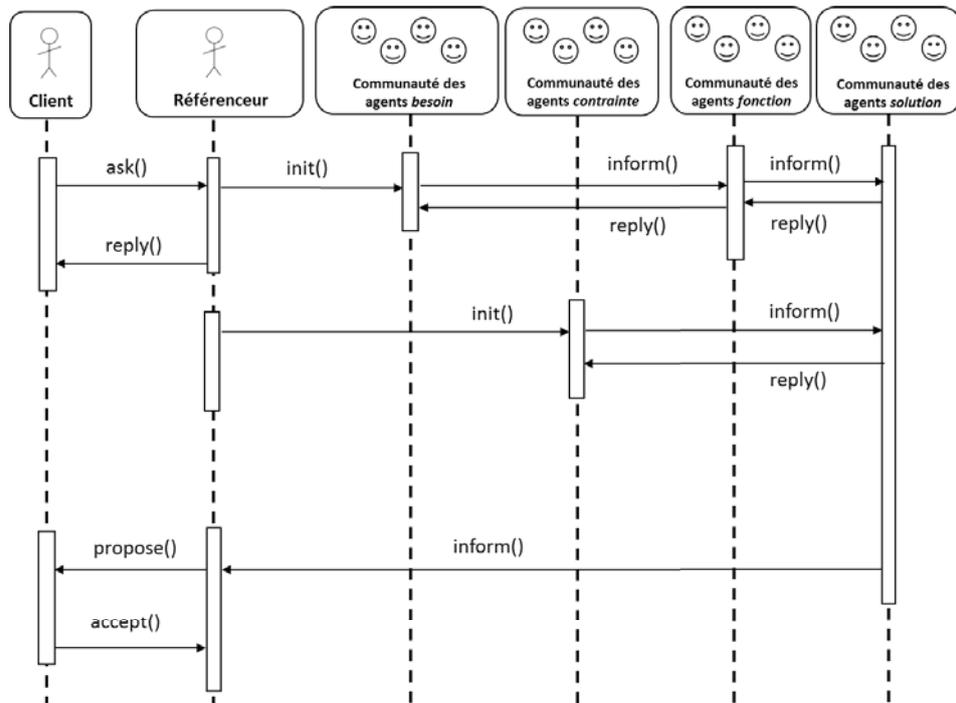


Figure 24. Diagramme de séquence représentant l'échange de messages entre agents et acteurs humains durant le processus de référencement

Les messages identifiés lors de ce scénario sont définis et répertoriés dans le tableau 4 qui suit :

Tableau 4. Table des messages

Émetteur	Message	Destinataire
Client (acteur humain)	ask(client,référenceur,y,S) ex : référencer (y) mon site S sur le mot-clé KW _i	Référenceur (acteur humain)
Référenceur (acteur humain)	init(r1,KW _i) ex : ajouter le nouveau mot-clé KW _i	Agent besoin (r1)
Référenceur (acteur humain)	init(c1,P _i) ex : ajouter la nouvelle page P _i	Agent contrainte (c1)
Agent besoin (r1)	inform(r1,f1,x,KW _i) ex : informer l'agent agent de créer (x) le mot-clé KW _i	Agent fonction (f1)
Agent contrainte (c1)	inform(c1,s1,x,P _i) ex : informer l'agent solution de créer (x) la page P _i	Agent solution (s1)
Agent fonction (f1)	inform(f1,s1,x,[P _i ,KW _i]) ex : informer l'agent solution de créer (x) le mot-clé KW _i sur la page P _i	Agent solution (s1)

La modélisation du processus de référencement nous a permis de mettre en évidence le rôle décisif joué par les agents *solution*. À partir des besoins et des contraintes identifiés, ils sont capables de proposer des solutions adéquates au référenceur qui pourra ensuite décider de les implémenter ou non avec l'accord du client. Les agents *solution* agissent en fonction de leurs connaissances, mais ils obéissent aussi à des règles de décision. Ces dernières doivent être pertinentes et doivent être mises à jour régulièrement pour que les agents *solution* puissent agir efficacement.

5. Discussion et conclusion

Dans ce chapitre nous avons présenté le modèle d'agents permettant la modélisation du processus de référencement. Après avoir décrit le modèle, les acteurs et les caractéristiques des agents impliqués, nous avons pu mettre en évidence leurs interactions et identifier les messages échangés ainsi que leurs contenus.

Cette démarche de modélisation à base d'agents a pu démontrer son efficacité à de nombreuses reprises et reste aujourd'hui très utilisée de manière générale pour simuler des processus de conception [63]. Par ailleurs, les caractéristiques des agents (distribution, autonomie,

flexibilité, coopération) nous ont permis de répondre de manière adéquate à nos besoins en termes de modélisation et de résolution de problème distribué pour le référencement. C'est la raison pour laquelle nous avons choisi d'utiliser cette démarche. Cela nous a permis de répondre à notre deuxième hypothèse :

Hypothèse n°2 : « *Le processus de référencement est modélisable* »

D'autre part, la décomposition du processus à partir du méta-modèle nous a permis de mettre en évidence des groupes d'agents disposant d'interactions, de rôles et d'une organisation spécifique. Les agents *solution* offrent des perspectives intéressantes dans le support à l'activité de référencement. Mais le caractère complexe de cette activité nous impose de constamment remettre en question les connaissances et les règles de décision de ces agents pour qu'ils puissent rester efficaces dans le temps.

Cet axe de recherche sera développé dans le prochain chapitre où nous présenterons également comment nous nous sommes lancés dans la conception d'un système d'aide à la décision à partir du modèle d'agents proposé dans ce chapitre.

Chapitre 5. Conception d'un système d'aide à la décision pour le référencement web

1.	<i>Introduction</i>	88
2.	<i>Vers un système d'aide à la décision pour le référencement web</i>	89
2.1	Objectif de notre système d'aide à la décision.....	89
2.2	Caractéristiques de notre système d'aide à la décision	91
3.	<i>Un système pour la simulation du processus de référencement web</i>	93
3.1	Calcul du classement	94
3.2	Application.....	96
3.2.1	Définition du projet	96
3.2.2	Simulation.....	97
4.	<i>Notre système d'aide à la décision</i>	99
4.1	Architecture de notre système d'aide à la décision.....	99
4.2	Étape 1 : Identifier l'impact des critères.....	102
4.2.1	Identification de l'impact des critères indépendants	103
4.2.1.1	Critère KWrText.....	104
4.2.1.2	Critère KWrTitle	105
4.2.1.3	Critère KWrH1.....	105
4.2.1.4	Bilan des expérimentations conduites sur des critères indépendants	106
4.2.2	Identification de l'impact des critères combinés.....	107
4.2.2.1	Bilan des expérimentations conduites sur des critères combinés	109
4.3	Étape 2 : Définition des ensembles flous.....	109
4.3.1	Bilan sur la définition des ensembles flous.....	111
4.4	Étape 3 : Définition des règles de décision floues	112
4.4.1	Bilan sur la définition des règles floues.....	114
5.	<i>Présentation du système d'aide à la décision</i>	114
5.1	Essai et validation du système en conditions réelles	120
6.	<i>Discussion et conclusion</i>	123

1. Introduction

Dans le chapitre précédent, nous avons proposé une modélisation du processus de référencement à partir d'un modèle d'agents. Ainsi, nous nous sommes focalisés sur certains groupes d'agents, notamment les agents *solution*, qui concentrent la partie de raisonnement et permettent les prises de décision. Dès lors ils jouent un rôle central dans le processus de référencement.

En poursuivant sur cette voie, nous avons cherché à développer un système d'aide à la décision construit à partir de notre modèle

d'agents et permettant une meilleure maîtrise du processus de référencement par le couple référenceur/client. Ce système vise à procurer aux référenceurs un gain de temps et d'efficacité durant la phase de « proposition des actions ». Une série d'expérimentations conduites sur des critères réels du référencement nous ont permis de définir des règles de décision floues que nous avons utilisées pour construire notre système d'aide à la décision.

L'objectif de ce chapitre est de développer un système d'aide à la décision permettant une meilleure maîtrise du processus de référencement, mais aussi une collaboration plus efficace entre le référenceur et son client. Cela nous permettra de répondre à notre troisième hypothèse :

Hypothèse n°3 : « *Le processus de référencement est mieux maîtrisable grâce à l'utilisation d'un système d'aide à la décision* »

Ce chapitre est composé de quatre sections. La première section présente l'objectif et les caractéristiques de notre système d'aide à la décision. La deuxième section expose un premier système permettant de simuler le classement de pages web afin de mieux comprendre le processus de référencement. La troisième section présente l'architecture de notre système d'aide à la décision et nous exposons comment, à partir d'expérimentations conduites sur des critères, nous avons pu générer des règles de décision. Dans la quatrième section, nous présentons l'interface de notre système d'aide à la décision ainsi que les résultats des tests effectués en conditions réelles. Nous terminons ce chapitre par une discussion et une conclusion.

2. Vers un système d'aide à la décision pour le référencement web

2.1 Objectif de notre système d'aide à la décision

Comme nous l'avons exposé dans le chapitre 3, durant la troisième étape du processus de référencement, le référenceur est amené à « proposer des actions » en adéquation avec les exigences et les besoins de son client pour améliorer le classement de son site web. Cette phase de travail, appelée « audit SEO » dans notre entreprise pilote, exige la réalisation d'un diagnostic complet du site web. Pour ce faire, le référenceur doit **utiliser de nombreux outils** que nous

avons présentés au chapitre 1, comme : le suivi de positionnement pour visualiser la position des mots-clés avant le début du travail, l'audit HTML pour vérifier quelles sont les balises à corriger, l'analyse de l'ancienneté pour vérifier si le site web possède un historique suffisant, etc. Par ailleurs, le référenceur va également **travailler de façon intuitive** [4] et faire confiance à son expérience, son savoir et son savoir-faire. Dans ce travail il n'est pas réellement guidé ou conseillé ce qui rend ce travail de **diagnostic très artisanal**. De plus, nous avons vu qu'un référenceur ne pourra jamais optimiser les centaines de critères utilisés par les moteurs de recherche pour classer un site, car il doit répondre à des **impératifs de temps et de coût**. Il doit donc être capable d'**analyser rapidement et efficacement** le site web du client afin de pouvoir **identifier et prioriser les actions** correctives à mettre en place.

Nous pensons que le référenceur a besoin d'être accompagné durant cette étape afin de pouvoir établir un diagnostic rapide et cohérent. Il en est même demandeur en accord avec notre enquête conduite dans la section 5.1. Dès lors, la conception d'un système d'aide à la décision permettant d'identifier rapidement et convenablement les critères à améliorer nous a semblé nécessaire. Ainsi le référenceur pourra plus facilement décider, en connaissances de cause, quelles sont les actions correctives à privilégier. Le système d'aide à la décision que nous avons conçu répond à cet objectif et permet au référenceur de gagner du temps et de l'efficacité durant cette phase de travail. Mais il a aussi pour vocation d'améliorer l'aspect collaboratif du travail de référencement. Selon Hoc [54], un référentiel commun permet d'améliorer la collaboration entre des personnes dont le niveau de connaissance est différent. Ainsi, il était nécessaire de réaliser deux interfaces : une interface « expert » dédiée au référenceur, et une interface « standard » dédiée au client. La première interface doit être capable d'aider le référenceur dans son travail de proposition d'actions. La deuxième interface doit permettre au client de gagner en autonomie en lui procurant une visibilité sur l'évolution du classement de son site web en temps réel. Le référenceur, qui aura également accès à l'interface « standard », pourra ainsi échanger avec son client selon un référentiel commun.

2.2 Caractéristiques de notre système d'aide à la décision

L'idée d'être assisté par un système informatique pour la prise de décision remonte aux années 1960. Les premières réalisations et implémentations de système d'aide à la décision quant à elles sont apparues dans les années 1980 [74]. Depuis, ils ont été proposés dans des domaines aussi variés que le transport [75]–[77] ou l'évaluation dans le contexte du e-learning [78]–[80]. Les principales caractéristiques d'un système d'aide à la décision sont les suivantes :

- Faciliter le processus de décision
- Soutenir plutôt qu'automatiser les décisions
- Permettre de gagner du temps
- Pour les systèmes les plus évolués : proposer des solutions optimales

Ces caractéristiques correspondent bien à l'outil que nous souhaitons concevoir pour aider le référenceur à mieux maîtriser le processus de référencement. Nous souhaitons aussi que le référenceur reste l'unique décideur quant aux choix des actions correctives à implémenter, ainsi nous avons décidé de ne pas automatiser complètement les prises de décisions [81].

Nižetić [74] a divisé les systèmes d'aide à la décision (SAD) en trois types:

- **Les SAD passifs** : ils aident le processus de prise de décision, mais ne peuvent pas identifier des propositions de décision ou des solutions.
- **Les SAD actifs** : ils identifient des propositions de décision ou des solutions
- **Les SAD coopératifs** : ils permettent au décideur de modifier la proposition de décision émise par le système. Le processus est ensuite répété jusqu'à ce qu'une solution satisfaisante soit générée.

Nous considérons notre système d'aide à la décision dans sa version actuelle comme un « SAD passif ». En effet notre objectif n'est pas de proposer de solutions explicites au référenceur telle que : « *ajouter quatre répétitions de mots-clés sur la page d'accueil* », mais plutôt

de l'orienter dans sa prise de décision comme par exemple : « *le nombre de mots-clés sur la page d'accueil est insuffisant, il faut corriger ce point en priorité pour espérer améliorer le classement du site* ». Le référenceur reste ainsi le décideur final des actions correctives à implémenter, nous ne faisons qu'orienter son choix en lui procurant un appui décisionnel sur les critères à corriger en priorité [81].

La structure d'un système d'aide à la décision est composée d'une interface utilisateur, d'un moteur d'inférence associé à une base de connaissance ou d'un modèle en fonction de son type. Nous pouvons distinguer trois types de système d'aide à la décision, le SAD guidé par les *connaissances*, le SAD guidé par un *modèle* et le SAD *hybride* qui cumule les deux aspects : *modèle* et *connaissances*. La figure 25 ci-dessous inspirée de Nižetić [74] permet de distinguer ces trois types de systèmes d'aide à la décision :

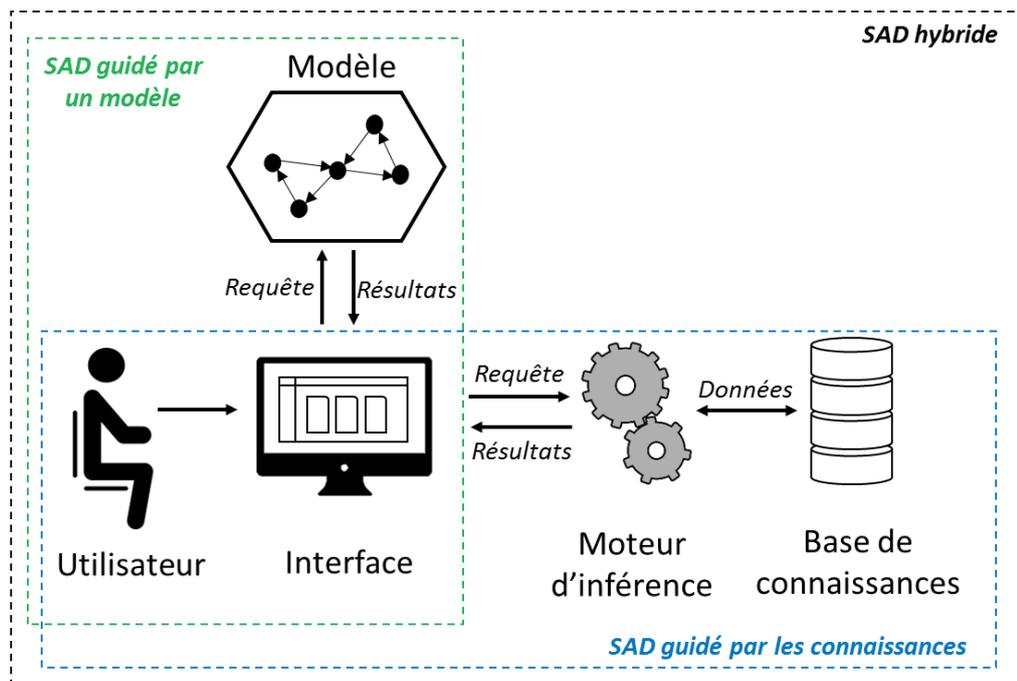


Figure 25. Trois types de systèmes d'aide à la décision : SAD guidé par un modèle, SAD guidé par les connaissances et SAD hybride

Notre système d'aide à la décision appartient à la catégorie des « SAD hybride » car il est basé sur un **modèle d'agents** où les agents *solution*, en fonction de leurs **connaissances** (règles de décision) vont pouvoir conseiller le référenceur dans le choix des actions à implémenter.

Dans le système que nous avons conçu, les agents *solution* jouent un rôle important, car c'est à leur niveau que se situent les prises de décision. Le paradigme agent a déjà été utilisé par plusieurs chercheurs [82] pour construire des systèmes d'aide à la décision. Nous avons considéré que ce type de démarche est bien adaptée à la résolution de notre problématique, notamment dans un contexte de collaboration référenceur/client [70], [83]. Notons que cette démarche a déjà été utilisée par d'autres chercheurs pour construire un système d'aide à la décision pour le référencement. Jankowski et al. [84] ont analysé des pages de sites web grâce à des agents dédiés au « *text mining* » (extraction de connaissances) et ils ont ensuite comparé leur classement. En identifiant des structures communes aux sites web bien classés, des agents dédiés à la « *clusterisation* » ont pu établir des groupements de pages en fonction de leurs caractéristiques et de leur classement. L'objectif était d'identifier les caractéristiques communes aux pages bien classées afin que l'utilisateur puisse mieux concevoir et mieux référencer son propre site. Comme nous le verrons dans la section 4, cette démarche diffère légèrement de la nôtre, néanmoins elle révèle le souhait et l'intention des chercheurs de bien vouloir aider le référenceur dans ses prises de décision.

3. Un système pour la simulation du processus de référencement web

Avant de concevoir notre système d'aide à la décision, nous nous sommes donnés comme premier objectif de concevoir un système permettant de simuler le **classement de pages selon une approche multi-critères** [85] à partir des connaissances de l'expert (référenceur). L'intérêt de simuler le fonctionnement d'un moteur de recherche est de pouvoir personnaliser la méthode de calcul du classement pour se rapprocher de la réalité tout en évitant de subir les inconvénients d'un vrai moteur de recherche, à savoir :

- La **concurrence** des sites web préexistants (dans notre système les sites web sont purement expérimentaux et peuvent être initialisés avec des valeurs de notre choix) ;
- Les **délais d'indexation** et de **prise en compte des modifications** souvent très longs (dans notre système quand un critère est modifié cela a une répercussion immédiate sur le classement du site web).

Sur la base d'une part des méthodes de travail des référenceurs mises en place dans notre entreprise pilote, et d'autre part des travaux de la littérature sur le sujet [7], [8], [19], [21]–[23], [28] nous avons identifié 15 critères pouvant influencer significativement le classement d'une page web. L'ensemble de ces critères est énoncé ci-dessous (1) :

$$C = \left\{ \begin{array}{l} c_1 = \text{âge_nom_domaine}; c_2 = \text{densité_mots-clés}; c_3 = \text{taux_rebond}; c_4 = \text{temps_passé}; \\ c_5 = \text{nombre_Google} + 1; c_6 = \text{nombre_Like_Facebook}; c_7 = \text{temps_chargement_page}; \\ c_8 = \text{longueur_<title>}; c_9 = \text{longueur_meta_description}; c_{10} = \text{répétition_mot-clés_<title>}; \\ c_{11} = \text{répétition_mots-clés_<h1>}; c_{12} = \text{nombre_backlinks}; c_{13} = \text{PageRank}; \\ c_{14} = \text{mot-clé_en_premier_<title>}; c_{15} = \text{mot-clé_en_premier_<h1>} \end{array} \right\} \quad (1)$$

3.1 Calcul du classement

Pour mettre en place notre simulation, nous avons identifié trois types d'influences : *positif*, *neutre* ou *négatif* que peut avoir un critère sur le classement d'une page web. Le tableau 5 donne des exemples d'intervalles *positif*, *neutre* et *négatif*, pour le critère $\{c_8 = \text{longueur de la balise <title>}\}$.

Tableau 5. Exemples de valeurs pour le critère $c_8 = \text{longueur de la balise <title>}$

Critères	Valeur positive (+)	Valeur neutre (0)	Valeur négative (-)
$\{c_8 = \text{longueur de la balise <title>}\}$	$a < x < b$ (a = 50, b=70)	$a < x < b$ (a=30, b=49)	$a < x < b$ (a=0, b=29) et (a=71, b=100)

Dans cet exemple, l'intervalle *positif* est compris entre $50 < x < 70$ caractères, l'intervalle *neutre* est compris entre $30 < x < 49$ caractères, et l'intervalle *négatif* est compris entre $0 < x < 29$ et $71 < x < 100$ caractères.

Ainsi chaque agent α_i peut calculer sa valeur *positive*, *neutre* ou *négative* pour chaque critère $c_j \in C$ et sa valeur correspondante x dans son domaine de définition. Trois opérateurs maximum, w_1 , w_2 and w_3 , associés respectivement aux mesures du *positif*, *neutre* et du *négatif* sont utilisés. Ils sont soumis à la contrainte suivante (2) :

$$\sum_{i=1}^n w_i = 1, 0 \leq w_i \leq 1, i = 1,2,3 \quad (2)$$

La figure 27 montre une représentation graphique de ces trois cas de figure. Nous pouvons ainsi identifier les trois intervalles : *négatif*, *neutre* et *positif* ainsi que leurs trois opérateurs respectifs : w_3 , w_2 et w_1 .

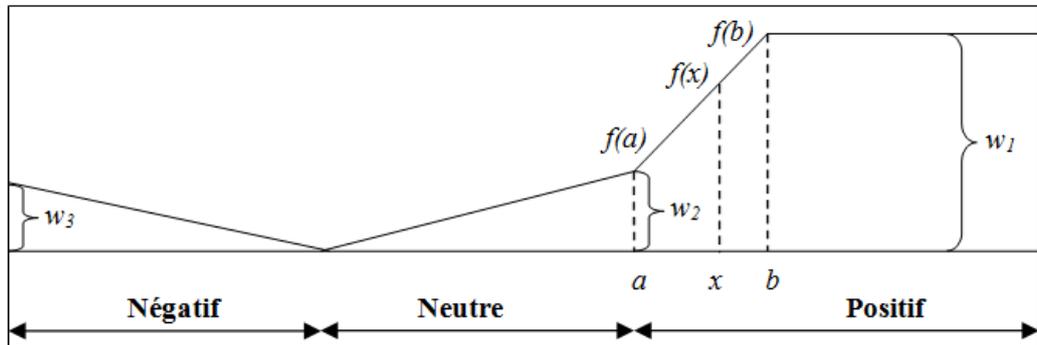


Figure 26. Représentation graphique pour nos trois cas de figure

Le tableau 6 expose les formules propres à chaque sous-domaine. Pour chaque sous domaine, il y a trois possibilités de formules en fonction de la valeur du critère :

Tableau 6. Équations proposées pour nos trois cas de figure

Cas	Positif (+)	Neutre (0)	Négatif (-)
Formules	$f(a) = w2$	$f(a) = 0$	$f(a) = w3$
	$f(b) = w1$	$f(b) = w2$	$f(b) = 0$
	$f(x) = w2 + [(w1-w2) \times (x-a)/(b-a)]$	$f(x) = w2 \times (x-a)/(b-a)$	$f(x) = w3 \times (b-x)/(b-a)$

La mesure de classement finale est définie comme suit (3) :

$$GlobalRankingMeasure(\alpha_i) = \sum_{\forall c_j \in C} Positive(c_j^i) + \sum_{\forall c_j \in C} Neutral(c_j^i) - \sum_{\forall c_j \in C} Negative(c_j^i) \quad (3)$$

où $GlobalRankingMeasure(\alpha_i)$ est la mesure finale du classement pour l'agent Page $\alpha_i \in A$. Par ailleurs, $Positive(c_j^i)$, $Neutral(c_j^i)$ and $Negative(c_j^i)$ sont respectivement les valeurs *positive*, *neutre* et *négative* pour le critère j pour l'agent Page α_i .

Pour résumer, en fonction du type d'influence que peut avoir un critère sur le classement (*positif*, *neutre* ou *négatif*) et de la valeur de ce critère, il faut appliquer une formule spécifique. En additionnant le résultat des formules pour tous les critères, il devient possible de déterminer le classement d'une page.

3.2 Application

Dans la section précédente, nous avons proposé une méthode de calcul permettant de simuler le classement de pages. Cette démarche a pour objectif de mieux comprendre le processus de référencement. Dans cette section nous allons mettre en pratique cette méthode grâce au développement d'un système de simulation de classement. Le test réalisé avec ce simulateur a été inspiré d'un projet réel conduit par notre entreprise pilote.

3.2.1 Définition du projet

Une pharmacie, que nous appellerons « *pharmacie F* » a contacté notre entreprise pilote pour améliorer le classement de son site sur le mot-clé « pharmacie + nom de la ville ». La pharmacie F était visible à la sixième position sur Google.fr, alors que les sites concurrents « A, B, C, D, E » présents sur le même mot-clé était respectivement à la 1^{ière}, 2^{ième}, 3^{ième}, 4^{ième}, et 5^{ième} position. En recherchant le terme « pharmacie + nom de la ville » sur Google.fr, plusieurs pages apparaissaient. Nous avons décidé d'ignorer dans le calcul du classement les pages qui n'étaient pas considérées comme des sites web de pharmacies (exemple : annuaires, pages jaunes, etc.) afin d'étudier des sites comparables. Nous avons aussi décidé de prendre en compte uniquement la page d'accueil, car cette page était présente pour chacun de nos six sites web sélectionnés. Ensuite nous avons identifié les valeurs des 15 critères pour chacun de ces sites en utilisant des outils de référencement tels que *SEOquake* (audit HTML), *Ahrefs* (analyse des liens retour), *GTMetrix* (analyse du temps de chargement), *SimilarWeb* (analyse de statistiques de visites) et *Whois* (analyse d'ancienneté). Tous les résultats obtenus ont été répertoriés dans le tableau 7 qui suit.

Tableau 7. Valeurs obtenues pour les critères de chacun des six sites de pharmacie sélectionnés

Page	Rang Google	c ₁ (mois)	c ₂ (%)	c ₃ (%)	c ₄ (s)	c ₅	c ₆	c ₇ (s)	c ₈	c ₉	c ₁₀	c ₁₁	c ₁₂	c ₁₃	c ₁₄	c ₁₅
A	#1	119	3.13	47	96	0	1	2.40	29	225	1	0	5	1	1	0
B	#2	81	3.89	50	132	0	0	4.50	96	143	1	1	3	1	0	1
C	#3	73	3.39	48	121	0	0	5.63	51	98	1	1	0	0	0	1
D	#4	37	1.58	51	164	0	1	6.84	33	0	1	1	1	0	1	1
E	#5	14	2.21	47	125	0	0	5.72	65	0	2	0	6	0	1	0
F	#6	10	2.91	52	153	0	0	3.74	69	0	2	0	2	0	0	0

3.2.2 Simulation

L'ensemble des agents *Page* ont été divisés en deux groupes : les agents *Page* pharmacie, et les agents *Page* non-pharmacie comme présenté sur la figure 27. Les agents *Page* pharmacie représentent les pages d'accueil des sites A, B, C, D, E et F. Les agents *Page* non-pharmacie représentent les pages avec lesquelles les agents *Page* pharmacie sont liés (avec un lien retour : *backlink*).

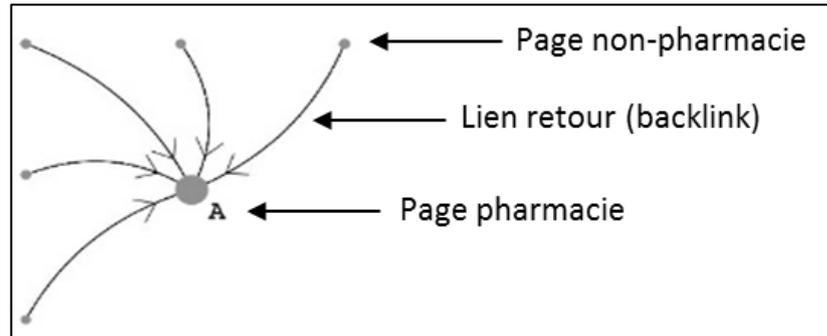


Figure 27. Différence entre Page pharmacie et Page non-pharmacie

Après avoir identifié les caractéristiques et les interactions entre les agents *solution*, nous avons implémenté notre modèle de simulation grâce à *NetLogo*³¹ qui est un environnement de programmation multi-agents. Notre simulation a été conçue comme un système de mesure de classement, indépendant du facteur temps afin de visualiser immédiatement le classement des pages. Quand tous les calculs sont terminés, la simulation prend fin. Dès lors, le score du classement de chaque page est calculé grâce à la *GlobalRankingMeasure* définie précédemment. Sur la figure 28 qui représente une capture d'écran de notre simulation, le classement est visible en bas à gauche dans la rubrique « *Ranking* ».

³¹ <https://ccl.northwestern.edu/netlogo/index.shtml>

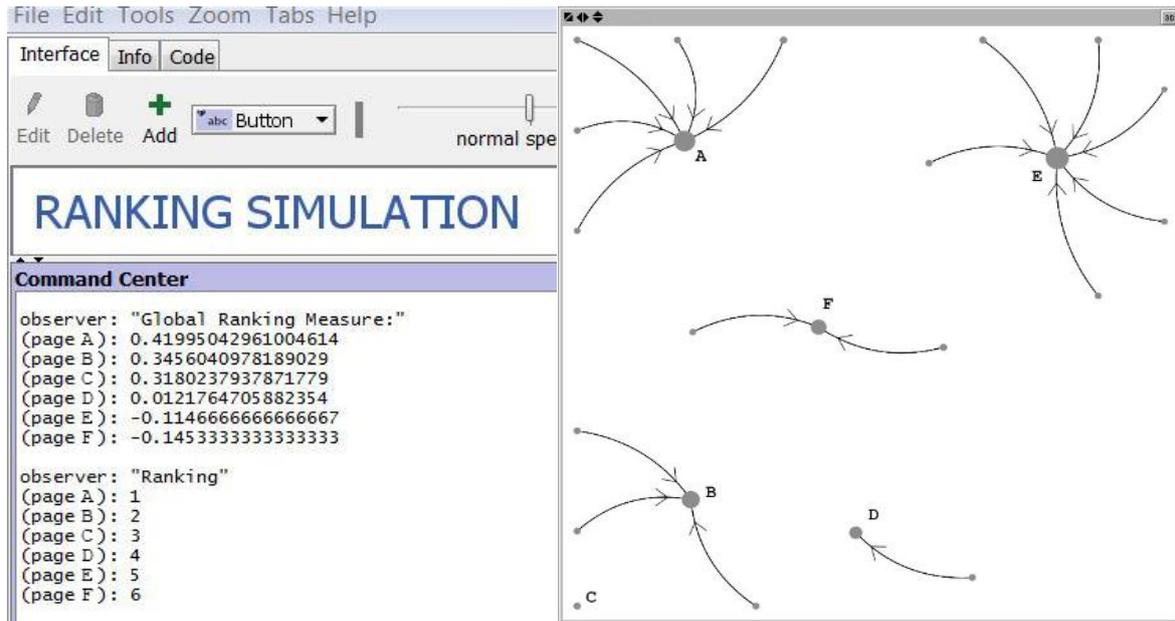


Figure 28. Simulateur de référencement multi-agents (environnement NetLogo), affichage du classement des six sites de pharmacies

Comme nous pouvons le voir sur la figure 28, le classement de nos six sites de pharmacies est similaire au classement établi par Google. Ainsi, cette simulation du processus de référencement nous a permis d'identifier dans un premier temps les critères ou les groupes de critères qui influencent le plus le processus de référencement comme « l'ancienneté du nom de domaine ». Par ailleurs, cette simulation nous a permis dans un deuxième temps de faire varier dynamiquement l'influence des critères sur le classement, comme pourrait le faire un vrai moteur de recherche. Enfin, à la manière d'un référencieur, nous avons aussi modifié les valeurs de ces critères pour faire varier le classement des sites web.

Cette première simulation nous a permis de mieux comprendre le fonctionnement d'un système de classement. Même si les moteurs de recherche communiquent peu sur le fonctionnement de leurs algorithmes de classement, il est possible de se rapprocher de leur mode de fonctionnement, voire même de répliquer leur classement comme nous avons pu le faire à travers notre simulation.

Pour améliorer davantage cette simulation, il faudrait pouvoir réaliser des tests à plus grande échelle, avec différents mots-clés et davantage de critères. Mais pour notre projet, cette simulation a été suffisante, car elle nous a permis de **mieux comprendre le fonctionnement du**

processus de référencement et de mettre à l'épreuve notre modèle d'agents dans le calcul du classement d'une page. Dans la section suivante, nous présentons le développement de notre système d'aide à la décision qui a pour objectif, comme déjà évoqué, de mieux maîtriser le processus de référencement.

4. Notre système d'aide à la décision

4.1 Architecture de notre système d'aide à la décision

L'architecture de notre système d'aide à la décision est basée sur le modèle d'agents présenté dans le chapitre précédent. La figure 29 ci-dessous est la représentation UML de notre modèle d'agent. Chaque agent possède des connaissances qui leur sont propres et ils ont la capacité d'envoyer des messages. Par ailleurs, ils appartiennent tous à une communauté d'agents. En vert nous avons identifié les agents *solution* qui concentrent la partie prise de décision.

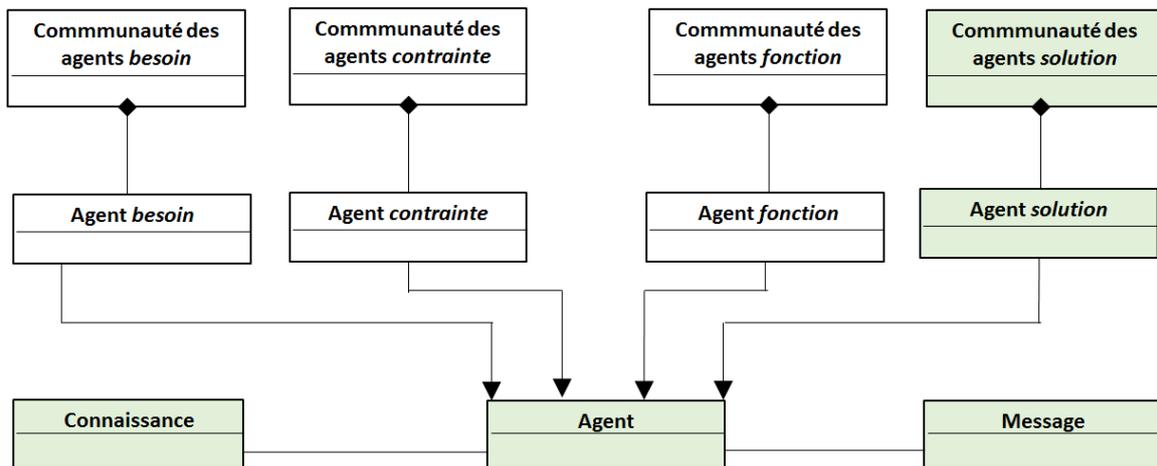


Figure 29. Diagramme de classes UML du processus de référencement

Le diagramme de classes UML sur la figure 30 ci-dessous nous expose l'architecture d'une page. Comme nous l'avons expliqué au chapitre précédent, une page est composée de critères de type « off-page » (ex : nombre de backlinks) ou « on-page » (ex : répétitions du mot-clé dans le texte).

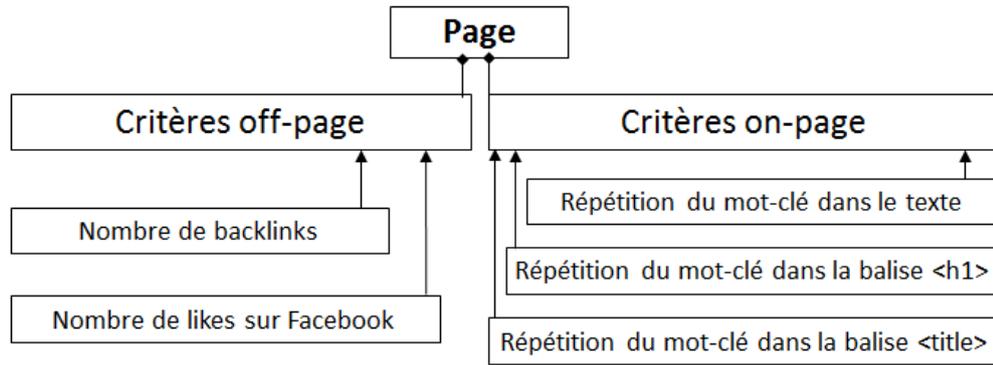


Figure 30. Diagramme de classes UML représentant l'agent solution Page

Dans notre système d'aide à la décision, les agents *Page* et *Critères* interagissent pour trouver une solution adaptée qu'ils suggèrent ensuite au référencieur. Le diagramme de séquence sur la figure 31 ci-dessous expose ces interactions. Dans notre exemple l'agent *Page*, en interagissant avec les agents *Critères*, est capable de calculer sa propre valeur et ainsi de proposer une action corrective pour améliorer son classement.

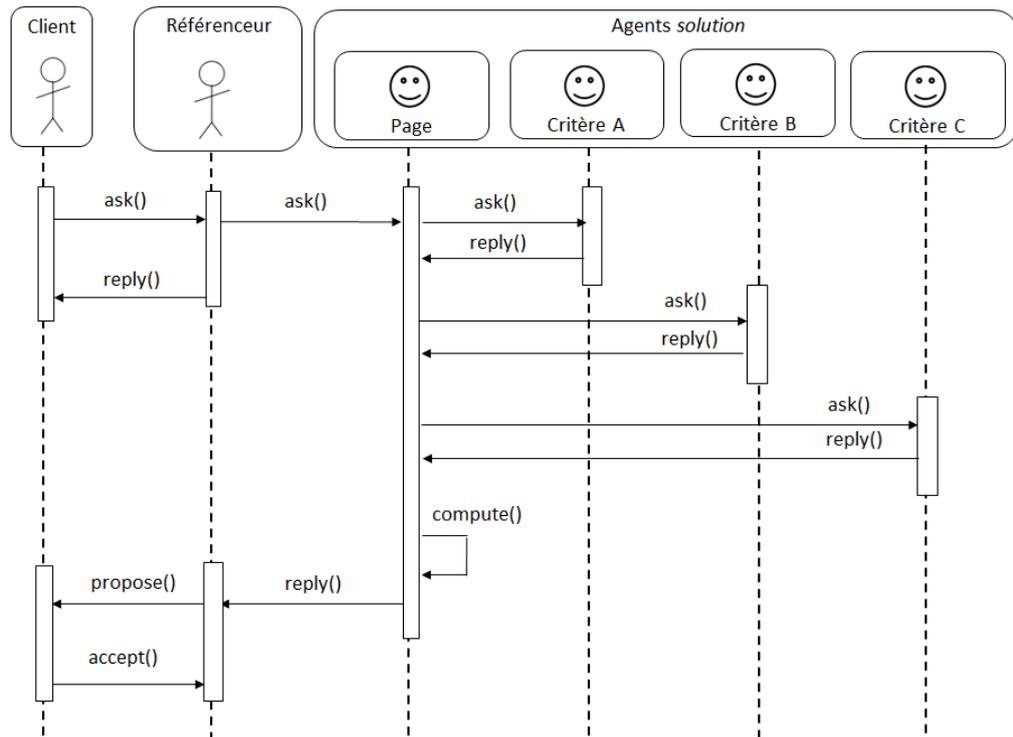


Figure 31. Diagramme de séquence représentant l'échange de messages entre agents solution et acteurs humains durant le processus de référencement

Pour soutenir, contester, voire orienter efficacement les choix du référencieur, les agents *solution* présents dans notre système sont ca-

pables d'identifier les critères qui pénalisent le classement d'une page web et peuvent prioriser les actions correctives à mettre en place. Pour cela, ils obéissent à des **règles de décision** intégrées au sein de leurs connaissances. Nous savons que dans le domaine du référencement, une décision peut aboutir à des **résultats incertains**. C'est la raison pour laquelle nous avons choisi de créer des règles en utilisant la logique floue.

La théorie des sous-ensembles flous permet de décrire des éléments dont les frontières sont mal définies [86], [87]. En effet, la notion d'ensemble flou est définie par l'appartenance graduelle d'un élément à une classe. Dès lors, un élément peut appartenir plus ou moins fortement à une classe, selon un degré d'appartenance variant entre 0 et 1. La logique floue se distingue donc de la logique booléenne classique où seulement deux états sont possibles : VRAI ou FAUX. En logique floue, plusieurs états sont possibles (ex : classement excellent, classement moyen, mauvais classement). Dans notre cas l'utilisation de la logique floue permet, contrairement à la logique booléenne classique, d'ajuster dynamiquement l'impact de nos critères afin de faire face à l'évolution constante des algorithmes de classement. Ainsi, pour chacun de nos critères, nous avons pu identifier trois ensembles flous (*bad, medium et excellent*) correspondant à l'influence que peut avoir un critère sur le classement d'un site web

Pour définir nos règles de décision floue, notre travail s'est décomposé en trois étapes :

Étape 1 : Identifier l'impact des critères (de manière indépendante puis combinée) grâce à des expérimentations conduites en conditions réelles

Étape 2 : Définir des ensembles flous correspondants à l'influence des critères sur le classement d'un site web (*bad, medium ou excellent*)

Étape 3 : Définir des règles de décision en fonction de l'impact des critères et des ensembles flous que nous aurons identifiés

Nous exposons chacune de ces étapes dans la section suivante.

4.2 Étape 1 : Identifier l'impact des critères

Pour que notre système puisse identifier et prioriser des actions correctives de manière cohérente, il nous a fallu tester l'impact de critères couramment utilisés par les référenceurs. Notre démarche se distingue de celle utilisée par Jankowski et Grzasko [84] puisque pour créer nos règles de décision nous n'avons pas souhaité analyser l'influence des critères sur des sites web préexistants. Les sites web préexistants possèdent des caractéristiques et des critères qui ne sont pas tous identiques, dès lors cela complique le travail d'identification. En effet, pour tester l'impact d'un critère précis, il ne faut pas que les critères non-testés perturbent la mesure. Pour tester nos critères, nous avons donc préféré créer nous-mêmes des sites web pour ensuite les indexer sur un moteur de recherche réel. Cette approche permet d'initialiser les valeurs des critères non-testés de manière identique afin d'éviter qu'ils n'interfèrent dans le classement des sites. Cette démarche de test de critères a déjà été utilisée par Zhang et Dimitroff [22], [23] et permet d'obtenir des résultats intéressants et pertinents. En étudiant les sites bien classés, nous avons ainsi pu identifier les valeurs optimales que devaient prendre les critères testés. Ces observations nous ont permis de définir des règles de décision que nous décrivons dans la section 4.4.

Pour mener à bien cette étude, nous avons choisi de nous concentrer sur trois critères fréquemment utilisés par les référenceurs au sein de notre entreprise pilote ; à savoir la répétition de mots-clés dans les balises <title>, la répétition de mots-clés dans les balises <h1> et enfin la répétition de mots-clés dans le contenu texte. Nous nous sommes demandé si ces répétitions de mots-clés avaient réellement un impact sur le classement d'un site web. Ainsi, notre panel est composé de trois critères que nous appellerons *KWrTitle*, *KWrH1* et *KWrText* :

- **KWrTitle** : nombre de répétitions d'un mot-clé dans la balise <title>
- **KWrH1** : nombre de répétitions d'un mot-clé dans la balise <h1>
- **KWrText** : nombre de répétitions d'un mot-clé dans le texte

Les valeurs testées pour ces critères sont indiquées ci-dessous. Ces valeurs de test ont été choisies en adéquation avec les méthodes de travail des référenceurs de notre entreprise pilote et des travaux de recherche de Zhang et Dimitroff [22], [23]. Selon ces chercheurs, au-delà d'un certain nombre de répétitions, le classement d'un site web peut diminuer (ex : un point de non-retour a été identifié à partir de quatre répétitions de mot-clé dans la balise <title>) :

- **KWrText** {0; 2; 4; 6; 8}
- **KWrTitle** {0; 1; 2; 3; 4}
- **KWrH1** {0; 1; 2; 3; 4}

4.2.1 Identification de l'impact des critères indépendants

Pour commencer notre étude, nous avons testé les critères de manière indépendante, car ces tests ont l'avantage d'être assez rapides à mettre en place. Pour tester nos trois critères, nous avons donc réalisé trois expérimentations. Nous avons choisi d'utiliser un vrai moteur de recherche pour tester nos critères afin que nos règles de décision puissent être cohérentes avec la réalité. Pour ce faire, nous avons choisi d'utiliser *Google.fr* car, comme vu au chapitre 1, c'est actuellement le moteur de recherche le plus utilisé en France. C'est également sur ce moteur de recherche que travaille la majorité des référenceurs dans notre entreprise pilote.

Pour réaliser ces trois expérimentations, nous avons créé trois groupes de sites web constitués d'une seule page web contenant chacune le même volume de texte (250 mots). Ce texte était généré automatiquement et aléatoirement³² de manière à ne pas créer de contenu dupliqué pour éviter les pénalités de la part du moteur de recherche. Par ailleurs, nous avons attribué une valeur identique aux critères non-testés (ex : âge du nom de domaine = 0, volume de texte = 250 mots, nombre de backlinks = 0, etc.) afin d'éviter qu'ils n'influencent le classement final. Chacun de ces sites web a été indexé puis mis en concurrence sur des mots-clés inexistantes (exemple : *topiletubizebuji*), en effet, sur ce type de requête, le moteur de re-

³² Outil de génération de texte aléatoire, <http://enneagon.org/phrases>

cherche ne renvoie aucun résultat. Ce choix était volontaire afin d'éviter que des sites préexistants n'interfèrent dans les résultats de ces tests.

4.2.1.1 Critère *KWrText*

Notre première expérimentation avait pour objectif de comparer de manière indépendante l'influence du critère *KWrText* (nombre de répétitions du mot-clé dans le texte) sur le classement d'un site web. Pour conduire cette expérimentation, nous avons utilisé 5 noms de domaines sur lesquels nous avons créé 5 sites web avec des valeurs distinctes pour le critère *KWrText*, à savoir {0 ; 2 ; 4 ; 6 ; 8} correspondant au nombre de répétitions de mots-clés dans le texte. Pour réaliser ce test, nous avons mis en concurrence ces 5 sites web sur un mot-clé inexistant pour éviter que des sites préexistants s'insèrent dans notre classement final. Ce test a permis d'obtenir un premier classement au **bout d'un seul jour** ce qui est vraiment très rapide. Suite à cela nous avons encore attendu un mois supplémentaire pour vérifier que le classement était bien stabilisé. Nous avons répliqué cette expérimentation une seconde fois en créant 5 nouveaux sites web sur un mot-clé différent, mais toujours inexistant. Le classement obtenu fut identique pour ces deux mêmes expérimentations, il est exposé ci-dessous :

- 1^{ière} position** : site avec 8 répétitions du mot-clé dans le texte
- 2^e position** : site avec 6 répétitions du mot-clé dans le texte
- 3^e position** : site avec 4 répétitions du mot-clé dans le texte
- 4^e position** : site avec 2 répétitions du mot-clé dans le texte
- Non classé** : site avec 0 répétition du mot-clé dans le texte

Les résultats révèlent que le site web possédant le **nombre maximum de répétitions** de mot-clé (8 répétitions) dans le texte obtient le meilleur classement. Ces résultats corroborent ceux obtenus par Zhang et Dimitroff [22], [23] car selon eux plus le mot-clé est répété dans le texte et meilleur sera le classement du site web. Ils n'ont constaté aucun point de non-retour, néanmoins nous déconseillons de sur-optimiser ce critère pour deux raisons. La première raison est qu'en répétant trop fréquemment le mot-clé dans le texte, cela peut perturber, voire empêcher la lecture de la page ; dans ce cas les internautes peuvent être amenés à quitter le site. Cet avis est partagé par les réfè-

renceurs de notre entreprise pilote et par les professionnels du référencement de manière générale [7], [8]. La deuxième raison est d'éviter une pénalité, en effet, comme vu au chapitre 1, la pratique du *keyword stuffing* (bourrage de mot-clé) est pénalisée par les moteurs de recherche [19].

4.2.1.2 Critère *KWrTitle*

Notre deuxième expérimentation avait pour objectif de comparer de manière indépendante l'influence du critère *KWrTitle* sur le classement d'un site web. Comme pour le test précédent, nous avons choisi un mot-clé inexistant pour éviter de perturber les résultats du classement. Par ailleurs, nous avons utilisé 5 noms de domaines sur lesquels nous avons créé 5 sites web avec des valeurs distinctes pour les critères *KWrTitle*, à savoir {0 ; 1 ; 2 ; 3 ; 4} correspondant au nombre de répétitions de mots-clés dans la balise <title>. Cette deuxième expérimentation nous a permis d'obtenir un premier classement **au bout d'un mois**. Nous avons encore attendu un mois supplémentaire pour vérifier que le classement était bien stabilisé. Nous avons répliqué cette même expérimentation deux fois et nous avons obtenu un classement identique. Les résultats de ce classement sont décrits ci-dessous :

- 1^{ère} position** : site avec 1 répétition du mot-clé dans la balise <title>
- 2^e position** : site avec 2 répétitions du mot-clé dans la balise <title>
- 3^e position** : site avec 3 répétitions du mot-clé dans la balise <title>
- 4^e position** : site avec 4 répétitions du mot-clé dans la balise <title>
- Non classé** : site avec 0 répétition du mot-clé dans la balise <title>

Les résultats révèlent que le site web possédant **une seule répétition** de mot-clé dans la balise <title> obtient le meilleur classement. Ces résultats diffèrent de ceux obtenus par Zhang et Dimitroff [22], [23] car ces chercheurs conseillent en effet de répéter trois fois le mot-clé dans la balise <title>. Cela confirme bien que les algorithmes de classement des moteurs de recherche sont en perpétuelle mutation [38].

4.2.1.3 Critère *KWrH1*

Nous avons mis en place une troisième expérimentation en utilisant un protocole identique pour le critère *KWrH1*. Nous avons testé ce

critère avec les valeurs {0 ; 1 ; 2 ; 3 ; 4} correspondant au nombre de répétitions de mots-clés dans la balise <h1>. Cette expérimentation nous a permis d'obtenir un premier classement au bout de **trois semaines**. Nous avons encore attendu un mois supplémentaire pour vérifier que le classement était bien stabilisé. Nous avons répliqué cette même expérimentation deux fois et le classement fut bien confirmé, il est décrit ci-dessous :

1^{ière} position : site avec 2 répétitions du mot-clé dans la balise <h1>

2^e position : site avec 4 répétitions du mot-clé dans la balise <h1>

3^e position : site avec 3 répétitions du mot-clé dans la balise <h1>

4^e position : site avec 1 répétition du mot-clé dans la balise <h1>

Non classé : site avec 0 répétition du mot-clé dans la balise <h1>

Les résultats obtenus nous ont permis de mettre en évidence que pour obtenir un bon classement, il fallait **répéter deux fois** le mot-clé dans la balise <h1>.

4.2.1.4 Bilan des expérimentations conduites sur des critères indépendants

Les résultats de ces différentes expérimentations nous ont permis d'identifier les valeurs optimales que doivent prendre nos critères afin d'obtenir un bon classement, à savoir pour résumer :

- 8 répétitions de mot-clé dans le contenu texte
- 1 répétition de mot-clé dans la balise <title>
- 2 répétitions de mot-clé dans la balise <h1>

Au cours de ces expérimentations, nous avons pu observer que les délais pour obtenir un premier classement étaient beaucoup plus longs avec les critères *KWrTitle* et *KWrHI* que pour le critère *KWrText*. Nous pouvons déduire de cette observation que le critère *KWrText* est un critère important pour le moteur de recherche Google. Pour indexer rapidement un site web sur ce moteur, le référenceur devra donc veiller à vérifier en priorité la présence du mot-clé à référencer dans le contenu texte.

4.2.2 Identification de l'impact des critères combinés

La deuxième partie de notre étude a été consacrée aux tests combinant plusieurs critères. Voici les deux couples de critères que nous avons testés :

- **Couple n°1** : *KWrText* {0 ; 2 ; 4 ; 6 ; 8} et *KWrTitle* {0 ; 1 ; 2 ; 3 ; 4}
- **Couple n°2** : *KWrText* {0 ; 2 ; 4 ; 6 ; 8} et *KWrH1* {0 ; 1 ; 2 ; 3 ; 4}

Pour réaliser ces expérimentations, nous avons dû créer deux groupes de 24 noms de domaines sur lesquels nous avons créé deux groupes de 24 sites web. Nous avons ainsi pu obtenir deux groupes de sites classés de la 1^{ière} position à la 24e position. La figure 32 qui suit montre l'évolution du classement des sites web pour le test du couple de critère n°2. Si les variations de classement sont fréquentes durant les premiers jours, celles-ci deviennent rares au bout de quelques semaines. Le classement définitif n'est donc pas instantané.

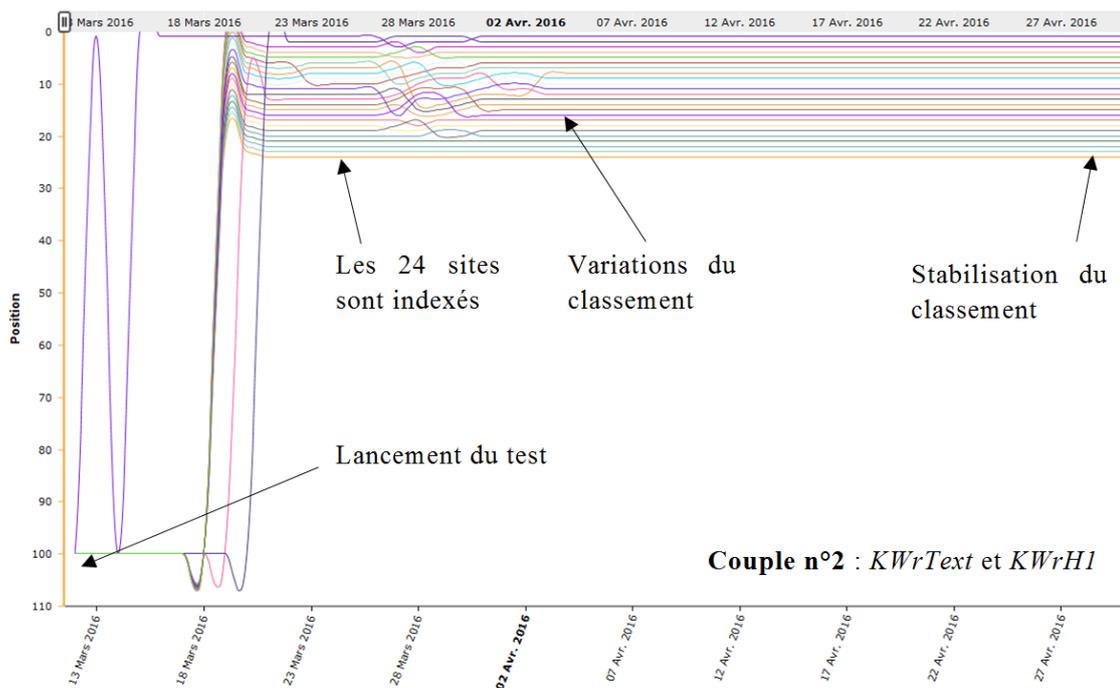


Figure 32. Évolution du classement au cours du temps pour le couple de critères n°2 : *KWrText* et *KWrH1*, interface Myposeo

Une fois le classement stabilisé, nous avons pu mesurer l'impact des critères testés. Les premiers résultats concernent le couple n°1 : *KWrText* et *KWrTitle*, ils sont présentés sur la figure 33. Les valeurs

des critères sont à chaque fois en abscisses et en ordonnées. Chaque point représente un site web, il y en a 24 au total. Les nombres de #1 à #24 permettent d'identifier le classement de chacun des sites web. Les zones « grises » permettent de repérer les sites ayant le meilleur classement (entre la première et la quatrième position sur la première page de résultat). Sur la figure 33, le site web classé en première position (#1) possède 8 répétitions du mot-clé dans le texte et 1 répétition du mot-clé dans la balise <title>. Le site web classé dernier (#24) possède 2 répétitions du mot-clé dans le texte et 0 répétition du mot-clé dans la balise <title>.

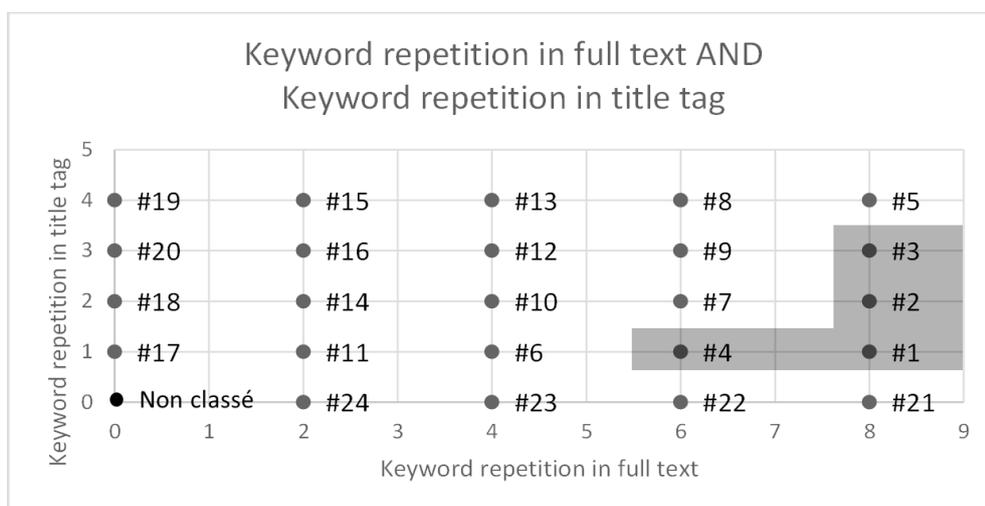


Figure 33. Classement stabilisé des 24 sites web sur le couple de critères n°1 : *KWrText* et *KWrTitle*

Les résultats pour le couple n°2 : *KWrText* et *KWrH1* sont présentés dans la figure 34. Le site classé en première position (#1) possède 8 répétitions du mot-clé dans le texte et 4 répétitions du mot-clé dans la balise <h1>. Le site classé en dernier (#24) possède 2 répétitions du mot-clé dans le texte et 0 répétition du mot-clé dans la balise <h1>.

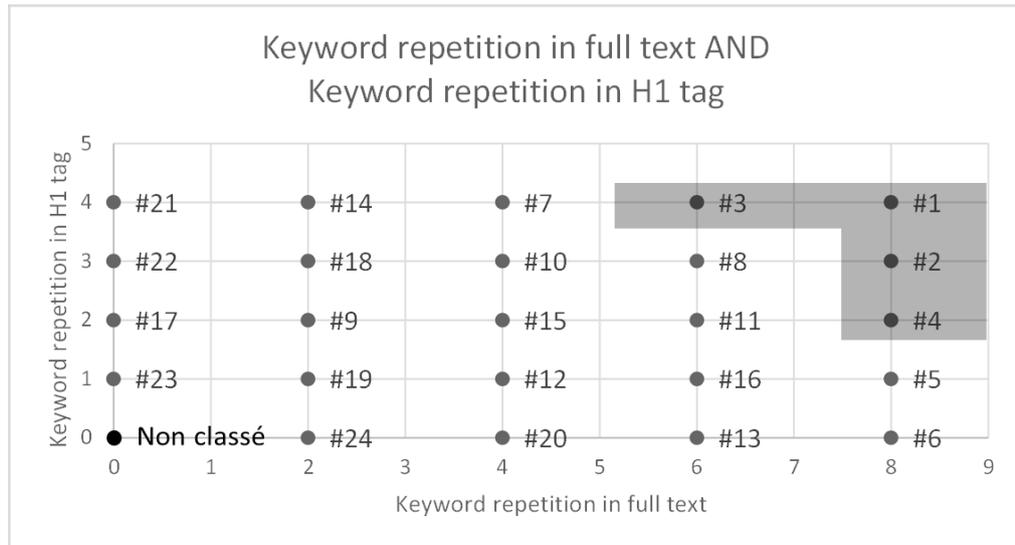


Figure 34. Classement stabilisé des 24 sites web sur le couple de critères n°2 : KWrText et KWrH1

Une analyse de variance à deux niveaux (*ANOVA2*) nous a permis de constater que le critère *KWrText* (nombre de répétitions de mots-clés dans le texte) avait statistiquement le plus fort impact sur le classement. (*Test couple n°1: F=0.68; P≤0.001 / Test couple n°2: F=0.91, P≤0.001*). Ainsi, **plus un mot-clé est répété dans le contenu texte** d'un site web et meilleur sera son classement. Néanmoins, l'impact du critère *KWrTitle* et *KWrH1* sur le classement n'est pas statistiquement représentatif (*Test couple n°1: F=0.28 / Test couple n°2: F=0.31, NS*).

4.2.2.1 Bilan des expérimentations conduites sur des critères combinés

Nous constatons à travers ces résultats que l'influence du critère *KWrText* sur le classement est très importante par rapport aux deux autres critères *KWrTitle* et *KWrH1*. En effet, parmi nos trois critères, c'est celui qui a le plus d'impact. Cela confirme aussi notre analyse précédente concernant les délais d'indexation très courts pour le critère *KWrText* : ce critère doit donc être corrigé en priorité par le référencement.

4.3 Étape 2 : Définition des ensembles flous

Les résultats des tests précédents nous ont permis de faire des observations concernant l'impact des critères *KWrText*, *KWrTitle*, et

KWrHI sur le classement de pages web. La deuxième étape est désormais de définir des ensembles flous. Pour rappel, ils sont au nombre de trois : *bad*, *medium* et *excellent*. Nous aurions pu définir davantage d'ensembles flous, mais dans nos exemples les intervalles de valeurs sont limités, dès lors, la définition de trois ensembles s'est avéré suffisamment représentative. Les ensembles flous ont été définis à partir des observations faites sur les résultats des expérimentations précédentes. Nous décrivons ci-dessous les ensembles flous que nous avons définis pour chacun de nos trois critères.

Pour le critère *KWrText*, la fonction « mauvais classement » (noté : *bad*) est comprise entre les valeurs $\{0 ; 3\}$, la fonction « classement moyen » (noté : *medium*) est comprise entre les valeurs $\{1 ; 7\}$ et la fonction « classement excellent » (noté : *excellent*) est comprise entre les valeurs $\{5 ; 8\}$. En utilisant le logiciel Matlab³³, nous avons pu représenter graphiquement ces ensembles flous, comme illustré sur la figure 35 ci-dessous :

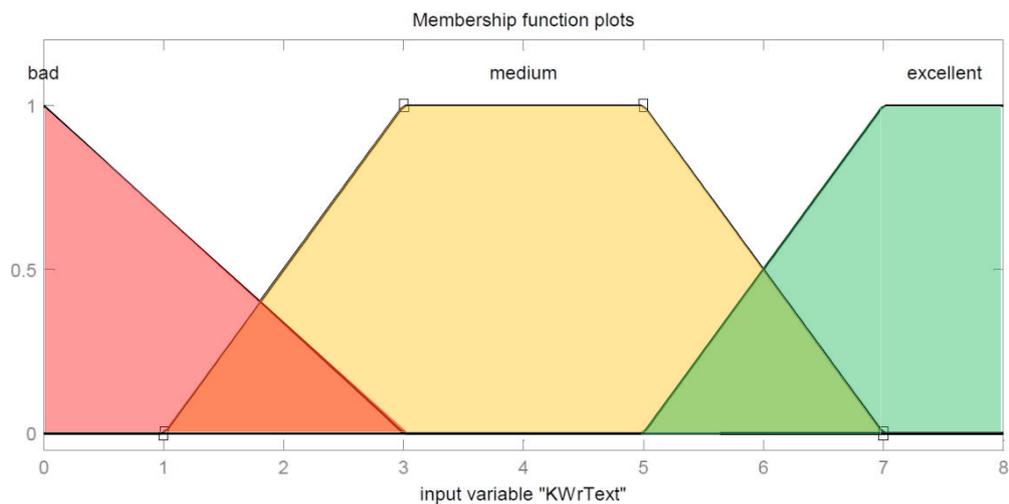


Figure 35. Trois ensembles flous pour le critère *KWrText* : *bad*, *medium* et *excellent*, Matlab

Pour le critère *KWrTitle*, la fonction « mauvais classement » est comprise entre $\{0\}$ et $\{4 ; 5\}$, la fonction « classement moyen » est com-

³³ Nous avons utilisé la boîte à outil Fuzzy Logic de Matlab

prise entre $\{2 ; 5\}$, la fonction « classement excellent » est comprise entre $\{1 ; 4\}$ comme le traduit la figure 36 :

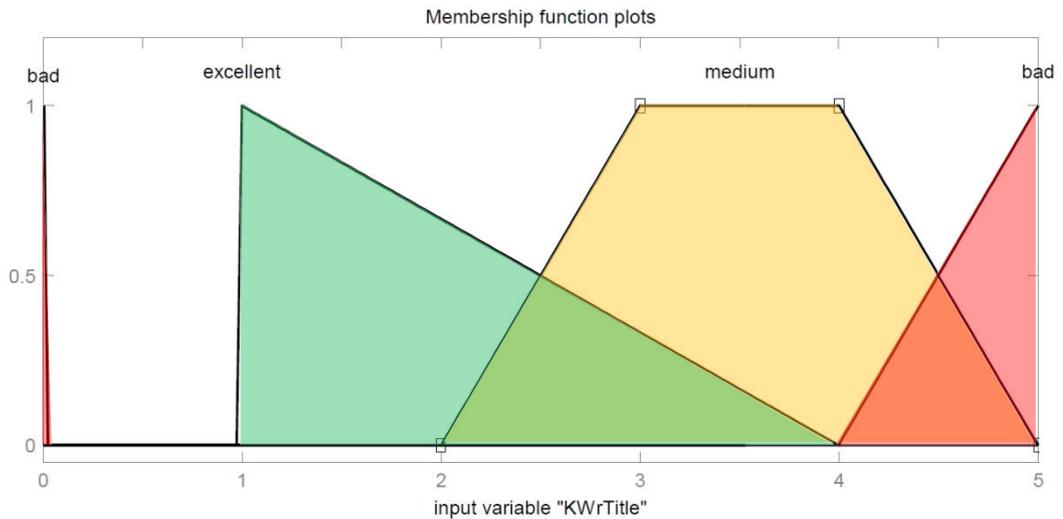


Figure 36. Trois ensembles flous pour le critère KWrTitle : bad, medium et excellent, Matlab

Pour le critère *KWrH1*, la fonction « mauvais classement » est comprise entre $\{0 ; 2\}$, la fonction « classement moyen » est comprise entre $\{2 ; 4\}$ et la fonction « classement excellent » est comprise entre $\{1 ; 4\}$ comme en témoigne la figure 37 :

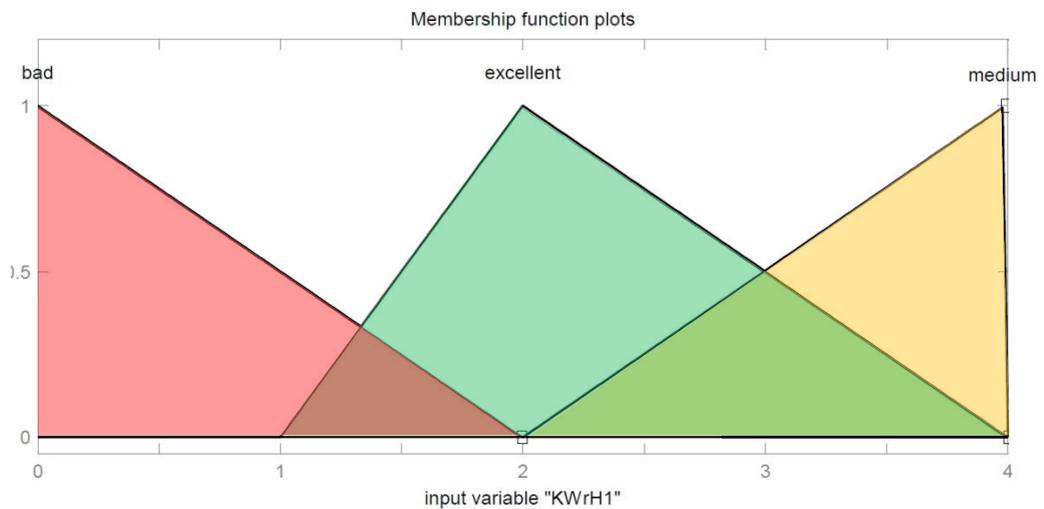


Figure 37. Trois ensembles flous pour le critère KWrH1 : bad, medium et excellent Matlab

4.3.1 Bilan sur la définition des ensembles flous

Les observations faites sur les résultats de nos expérimentations nous ont permis de définir trois ensembles flous pour chacun de nos cri-

tères. En calculant la fonction d'appartenance d'un critère à l'un de ces ensembles, il est possible de déduire son degré d'influence sur le classement d'une page web. Par ailleurs, nous pouvons à tout moment modifier ces ensembles flous pour rester en cohérence avec la réalité du terrain.

4.4 Étape 3 : Définition des règles de décision floues

Après avoir identifié trois ensembles flous pour chacun des critères testés, nous avons défini des règles de décision en accord avec nos observations sur les résultats des tests combinatoires. Comme présenté au chapitre 4, les règles de décision utilisées par les agents de notre modèle respectent le formalisme ECA :

[SI évènement(s) avec condition(s) ALORS action(s)]

Pour construire notre système d'aide à la décision, nous avons dû adapter ces règles à la logique floue, ce qui donne désormais :

SI {x EST A} alors {y EST B}

Voici quelques exemples de règles floues que nous avons pu définir sur la base des résultats de nos expérimentations :

```
IF (KWrText IS excellent) AND (KWrTitle IS excellent) THEN (Ranking IS excellent)
IF (KWrText IS excellent) AND (KWrTitle IS medium) THEN (Ranking IS excellent)
IF (KWrText IS excellent) AND (KWrTitle IS bad) THEN (Ranking IS bad)
```

Nous avons utilisé l'éditeur de règles floues de Matlab pour créer d'autres règles floues, comme traduit sur la figure 38 :

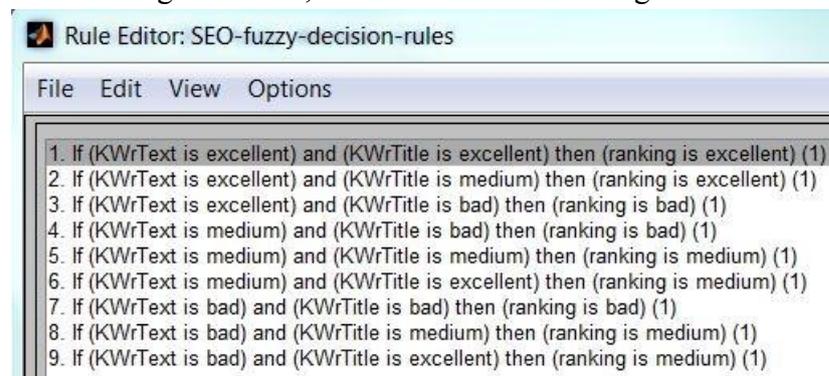


Figure 38. Règles floues pour les critères KWrText et KWrTitle, éditeur de règle MatLab

Les ensembles flous et les règles floues que nous avons définis peuvent être représentés graphiquement grâce à la fonction de Matlab qui permet de générer des graphiques de surface. Les figures 39 et 40 qui suivent illustrent bien ces deux représentations. La partie rouge représente les valeurs optimales des critères pour obtenir un bon classement. La figure 39 nous permet de déduire qu'un bon classement peut être obtenu à partir de 6 répétitions de mots-clés dans le texte et entre 1 et 4 répétitions de mot-clé dans la balise <title>.

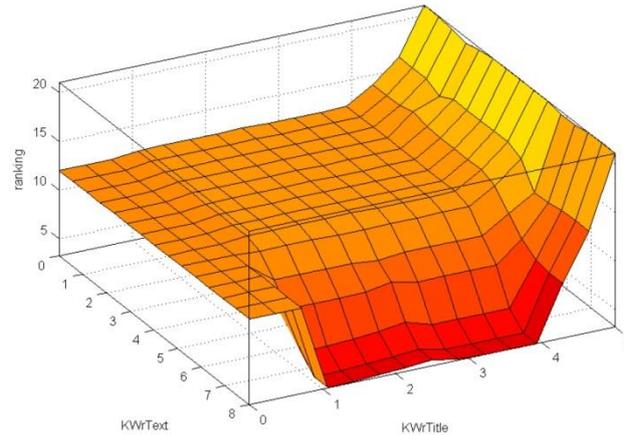


Figure 39. Graphique de surface du classement pour les critères KWrText et KWrTitle, Matlab

La figure 40 ci-dessous nous permet de déduire qu'un bon classement peut être obtenu à partir de 6 répétitions de mots-clés dans le texte et entre 1 et 4 répétitions dans la balise <h1>.

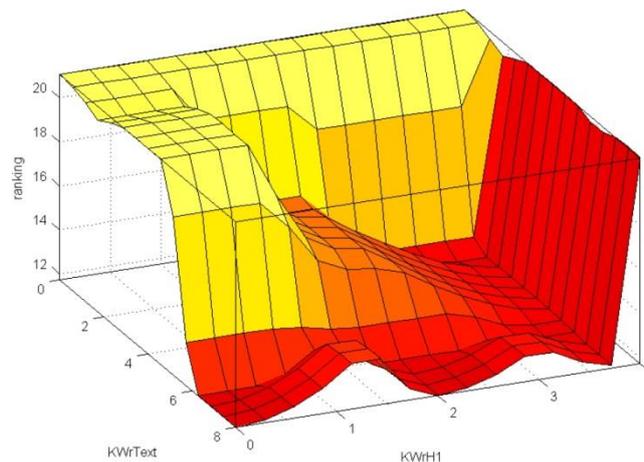


Figure 40. Graphique de surface du classement pour les critères KWrText et KWrH1, Matlab

Sur la base de nos observations et analyses issues de nos expérimentations conduites sur quelques critères, nous avons pu définir des en-

sembles flous et des règles de décision floues. Cela nous a permis de formaliser l'approche que nous proposons tout en offrant suffisamment de souplesse pour faire face aux mises à jour récurrentes des algorithmes de classement.

4.4.1 *Bilan sur la définition des règles floues*

La pertinence des règles floues est directement liée aux résultats, mais aussi à **la date des expérimentations**. La dimension temporelle est primordiale, il est donc nécessaire de renouveler régulièrement ce type d'expérimentations afin d'éviter que les règles ne deviennent obsolètes après la mise à jour des algorithmes de classement. Cependant, il convient aussi de préciser que ces expérimentations sont lourdes et longues à mettre en place. Par ailleurs, les résultats peuvent mettre plusieurs semaines voire des mois pour se stabiliser et ainsi devenir exploitables. Par ailleurs, pour être efficaces, ces expérimentations doivent impérativement combiner plusieurs critères. Dans notre recherche, nous avons combiné au maximum deux critères avec 5 valeurs ce qui exige néanmoins la création de 24 sites web différents ; pour combiner davantage de critères, le nombre de sites web à créer devient exponentiel. Par exemple, pour tester 3 critères sur 5 valeurs, il aurait fallu créer 124 sites web.

En conclusion, grâce à une série d'expérimentations réalisées sur des critères en condition réelle, nous avons pu définir des règles floues pertinentes et proches du contexte réel observé. Dans la prochaine section, nous intégrerons ces règles floues dans le système d'aide à la décision que nous avons conçu pour aider le référenceur et le client à mieux maîtriser le processus de référencement.

5. Présentation du système d'aide à la décision

L'architecture de notre système d'aide à la décision est composée des éléments suivants :

- **Un serveur** : il récupère (une fois par jour) les valeurs des critères de chaque site web puis les enregistre dans la base de données : « Critères » ;
- **Un moteur SEO** : il traite les données issues de la base de données « Critères » en fonction de règles de décision floues provenant de la base de données « Règles de décision » ;

- **Un client** : constitué de deux interfaces (expert et standard), il permet d'afficher les données traitées par le moteur SEO.

La figure 41 ci-dessous illustre la composition de notre système d'aide à la décision. Nous retrouvons le serveur, le moteur SEO, les deux bases de données (*Critères* et *Règles de décision*) et le client qui est constitué de deux interfaces (expert et standard) :

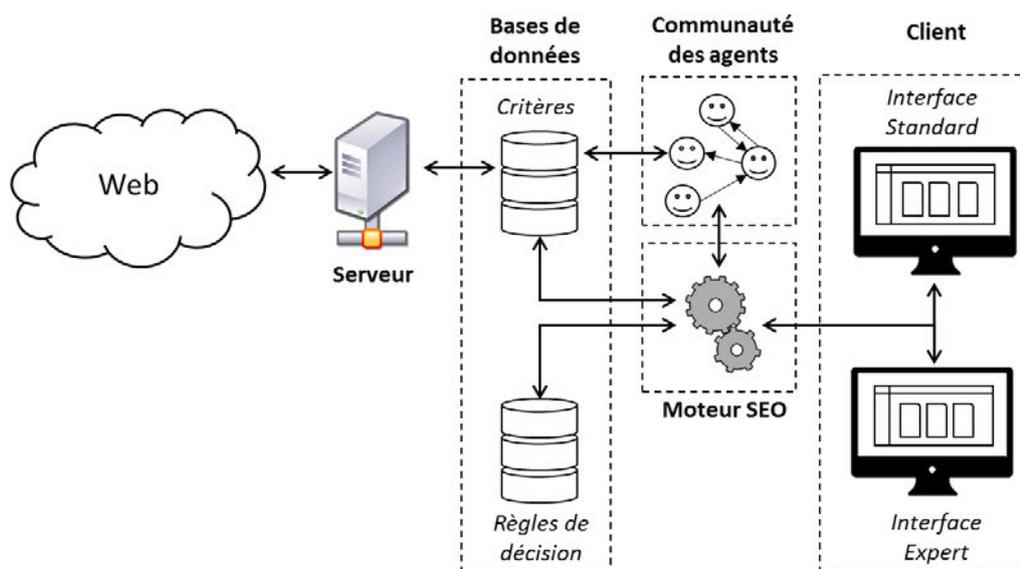


Figure 41. Architecture du système d'aide à la décision pour le référencement

Pour pouvoir proposer un diagnostic cohérent au référencier, notre système d'aide à la décision doit dans un premier temps récupérer les valeurs des critères de chaque site. Pour collecter ces valeurs, notre système utilise plusieurs outils présentés dans le tableau 8 qui suit :

Tableau 8. Outils utilisés par le système d'aide à la décision

Outils	Objectifs
Parseur HTML	Récupérer les valeurs des critères liés aux répétitions de mots-clés ; Détecter le type de plateforme CMS utilisée
MyPoseo	Connecté en API, cet outil permet de récupérer la liste des sites web des clients, les mots-clés associés à chaque client, ainsi que leur position
Google Analytics	Récupérer les valeurs des critères liés au taux de rebond, la durée passée sur le site et le temps de chargement des pages
Whois	Récupérer les valeurs d'ancienneté du site
Seo-kicks.fr/backlinks	Récupérer le nombre de backlinks et les domaines référents dont sont issus les backlinks
API Facebook/API Google+	Récupérer le nombre de <i>Like</i> sur Facebook et le nombre de <i>+1</i> sur Google+

Quand le système a connaissance des valeurs de chaque critère, il leur associe l'un de trois ensembles flous suivant : *bad*, *medium* et *excellent*, définis en fonction des résultats obtenus lors de nos expérimentations. Ces ensembles flous sont définis au format .fcl³⁴ dans notre système. La figure 42 ci-dessous expose les ensembles flous pour les critères *KWrText*, *KWrH1* et *KWrTitle* :

```

1 FUZZIFY KWrText
2     TERM bad := (0, 1) (3, 0);
3     TERM medium := (1, 0) (3, 1) (5, 1) (7, 0);
4     TERM excellent := (5, 0) (7, 1) (8, 1);
5 END_FUZZIFY
6
7 FUZZIFY KWrH1
8     TERM bad := (0, 1) (2, 0);
9     TERM medium := (2, 0) (4, 1) (4.01, 0);
10    TERM excellent := (1, 0) (2, 1) (4, 0);
11 END_FUZZIFY
12
13 FUZZIFY KWrTitle
14    TERM bad := (0, 1) (0.01, 0) (4, 0) (5, 1);
15    TERM medium := (2, 0) (3, 1) (4, 1) (5, 0);
16    TERM excellent := (0.99, 0) (1, 1) (4, 0);
17 END_FUZZIFY

```

Figure 42. Définition des ensembles flous pour les critères *KWrText*, *KWrH1* et *KWrTitle* au format .fcl

Une fois que le système a connaissance des ensembles flous auxquels appartiennent les critères, il applique les règles de décision. Ces règles de décision floues sont définies en fonction des résultats de nos expérimentations. Elles sont aussi écrites au format .fcl. La figure 43 ci-dessous expose quelques-unes de ces règles de décision :

³⁴ FCL (Fuzzy Control Language) est un langage pour implémenter la logique floue

```

332 RULE 0: IF (KWrText IS excellent) AND (KWrTitle IS excellent) THEN (Ranking IS excellent);
333 RULE 1: IF (KWrText IS excellent) AND (KWrTitle IS medium) THEN (Ranking IS excellent);
334 RULE 2: IF (KWrText IS excellent) AND (KWrTitle IS bad) THEN (Ranking IS bad);
335 RULE 3: IF (KWrText IS medium) AND (KWrTitle IS bad) THEN (Ranking IS bad);
336 RULE 4: IF (KWrText IS medium) AND (KWrTitle IS medium) THEN (Ranking IS medium);
337 RULE 5: IF (KWrText IS medium) AND (KWrTitle IS excellent) THEN (Ranking IS medium);
338 RULE 6: IF (KWrText IS bad) AND (KWrTitle IS bad) THEN (Ranking IS bad);
339 RULE 7: IF (KWrText IS bad) AND (KWrTitle IS medium) THEN (Ranking IS medium);
340 RULE 8: IF (KWrText IS bad) AND (KWrTitle IS excellent) THEN (Ranking IS medium);
    
```

Figure 43. Définition des règles de décision floues pour les critères KWrText et KWrTitle

Une fois que les données sont récupérées, stockées et traitées, notre système est capable de proposer un diagnostic via l'interface de diagnostic dédiée au référencier. La figure 44 expose l'interface « expert » dédiée au référencier :

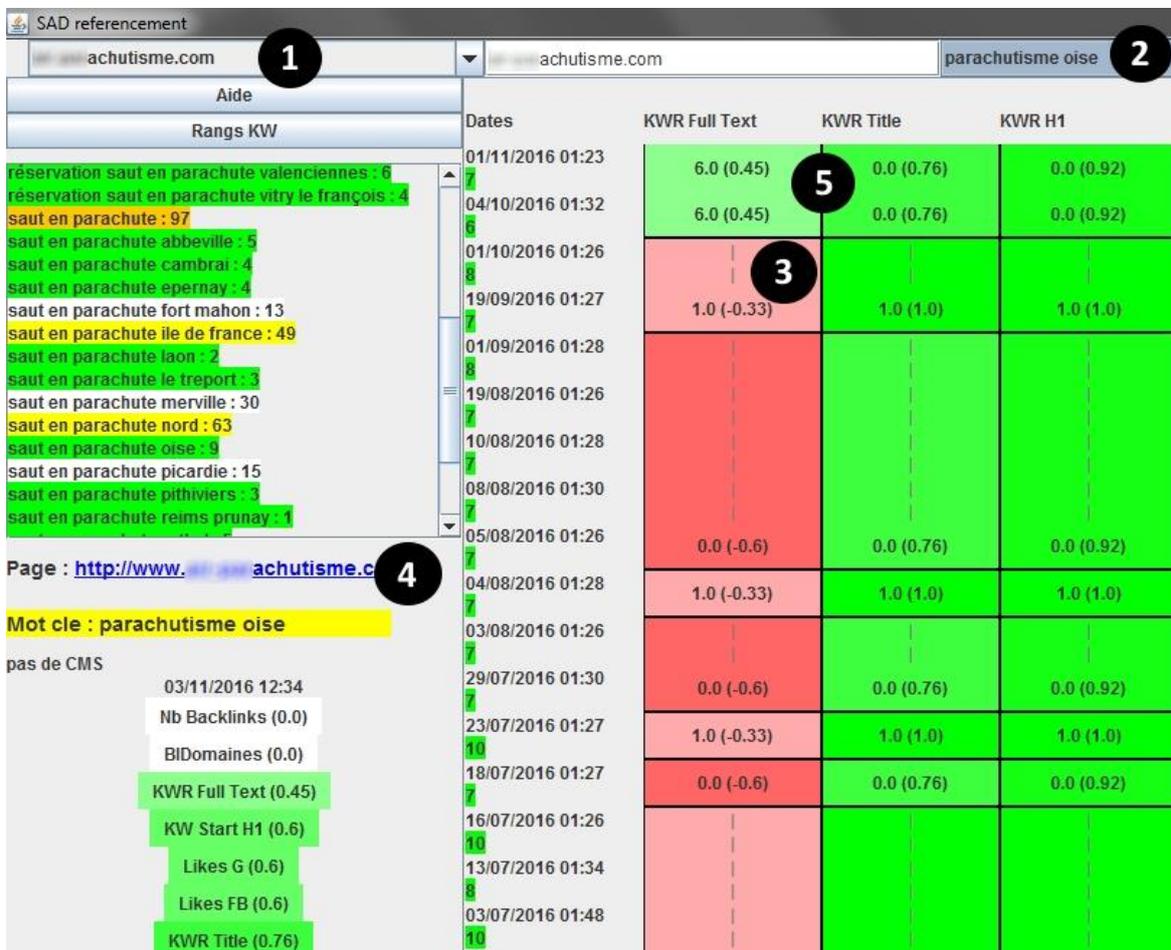


Figure 44. Interface expert du système d'aide à la décision

Sur la figure ci-dessus nous avons réalisé à titre indicatif le diagnostic d'un site web dédié au parachutisme (1). Parmi les mots-clés du

client, nous avons sélectionné le mot-clé « parachutisme oise » (2) qui était positionné à la 7e position (position visible en dessous du champ date) au 01/11/2016. Le diagnostic réalisé par le système montre que jusqu'au 01/10/2016 le critère à améliorer était *KWrText* (3) identifié en rouge dans la colonne. En effet sur la page concernée (4), il n'y avait pas assez de répétitions du mot-clé « parachutisme oise » dans le contenu texte. À noter que le dernier diagnostic effectué au 01/11/2016, indique que les critères *KWrText*, *KWrTitle* et *KWrHI* sont tous corrects (5), en effet, ils sont identifiés en vert dans leurs colonnes respectives. Nous pouvons en déduire que le référencier est intervenu sur le critère *KWrText*, puisque la ligne correspondante est passée du rouge au vert entre le 01/10/2016 et le 04/10/2016. Le système d'aide à la décision vérifie chaque jour les valeurs des critères. Si un changement est détecté, une nouvelle ligne de diagnostic est créée. Ainsi, si un critère a été corrigé par le référencier, la ligne qui était rouge au départ deviendra verte.

L'interface « standard » dédiée au client est présentée sur la figure 45. Sur cette interface le client peut suivre l'évolution du classement (1) de ses différents mots-clés (2). Sur l'exemple présenté, on remarque que le site web était positionné à la 11^e position sur le mot-clé « parachutisme oise » au 4 juin, soit la deuxième page de résultats ce qui n'est pas un bon classement. Au 15 octobre, le site était présent à la 8e position, soit la première page de résultats.

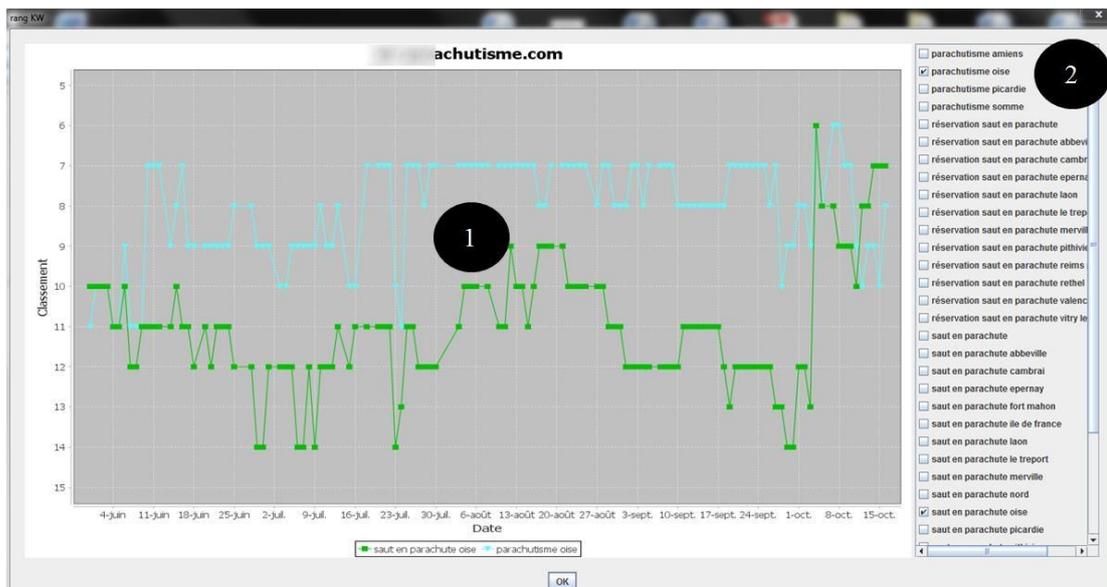


Figure 45. Interface standard du système d'aide à la décision

Suite à cette première version de notre système d'aide à la décision, nous avons décidé d'améliorer l'ergonomie de l'interface. La figure 46 expose cette deuxième version :

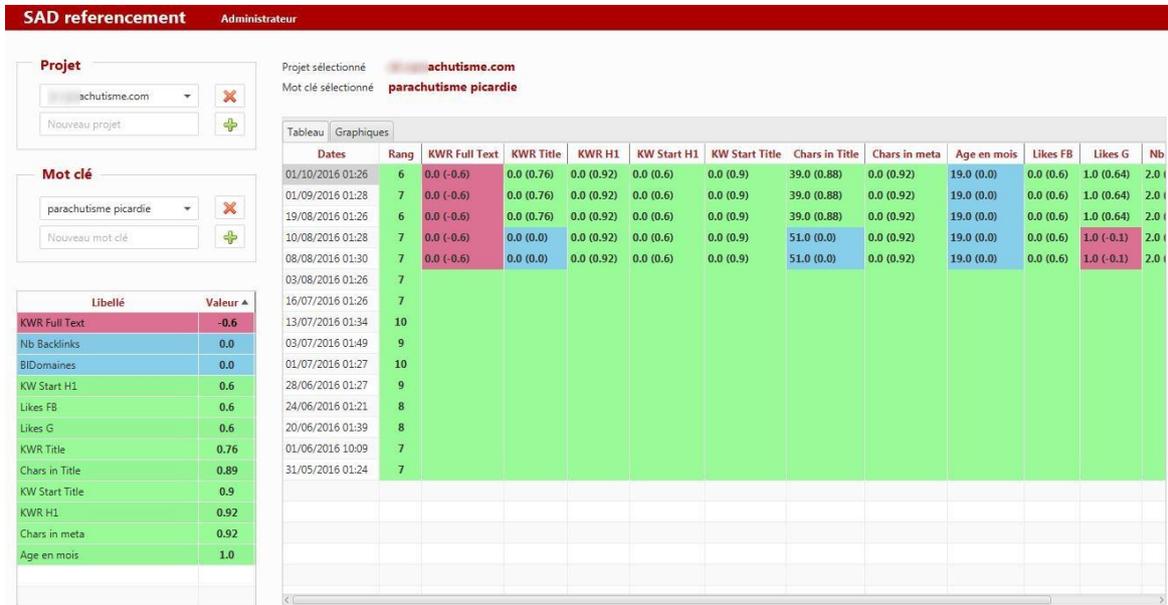


Figure 46. Deuxième version du système d'aide à la décision, interface expert

Nous avons également amélioré l'interface de la version client, la figure 47 illustre cette seconde version :

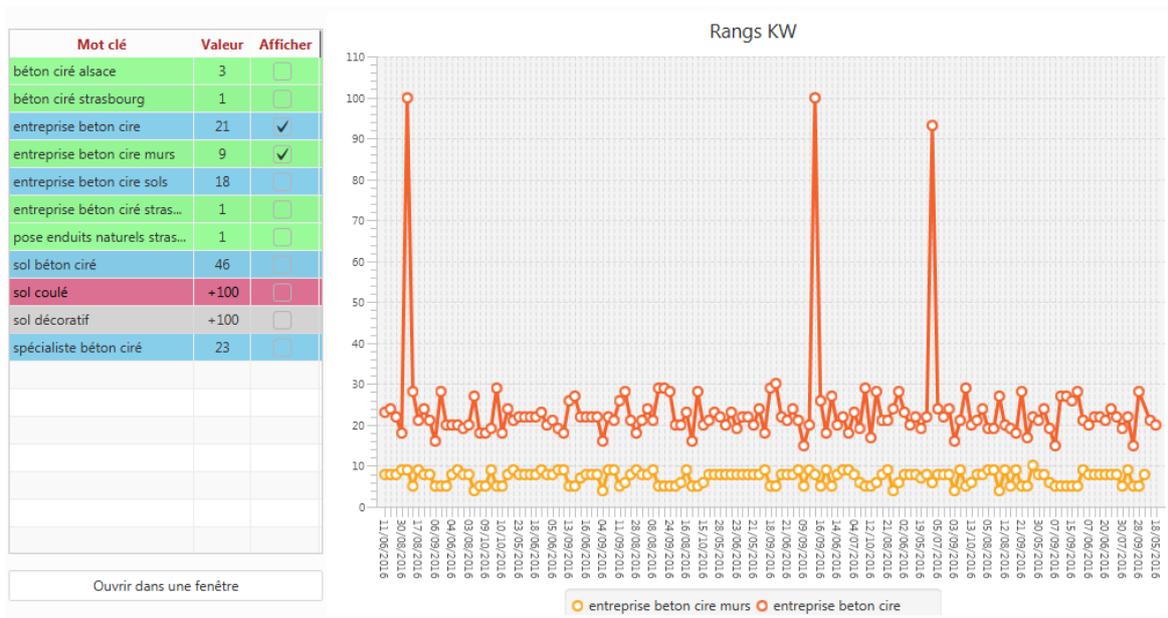


Figure 47. Deuxième version du système d'aide à la décision, interface standard

5.1 Essai et validation du système en conditions réelles

Après le développement de notre système d'aide à la décision, nous avons souhaité le faire valider auprès des trois référenceurs de notre entreprise pilote ; pour cela, nous avons réalisé une pré-étude. Un questionnaire fut ainsi distribué avant et après le test du système en condition réelle. Le premier questionnaire nous a permis de connaître les attentes des référenceurs avant le test du système. Nous avons répertorié les réponses à ce premier questionnaire dans le tableau 9 ci-dessous :

Tableau 9. Premier questionnaire

Questions	Réponses des trois référenceurs
A. Quels problèmes rencontrez-vous pour réaliser le diagnostic SEO d'un site ?	<ul style="list-style-type: none"> ➤ Perte de temps ➤ Beaucoup d'outils différents à utiliser
B. Quels sont les avantages de votre méthode de travail actuelle ?	<ul style="list-style-type: none"> ➤ Jugement humain pour réaliser le chiffrage (devis) ➤ Pertinence et fiabilité des résultats
C. Avez-vous besoin d'un système d'aide à la décision ?	<ul style="list-style-type: none"> ➤ Oui
D. Qu'attendez-vous d'un système d'aide à la décision coopératif pour le référencement web ?	<ul style="list-style-type: none"> ➤ Des indications/conseils sur les critères à optimiser ➤ Des indications/conseils sur les critères à ne pas optimiser ➤ Ajouter/supprimer des mots-clés à étudier ➤ Connaître la position du/des mot(s)-clé(s) à l'instant t ➤ Connaître l'évolution de la position du/des mot(s)-clé(s) ➤ Que le client puisse suivre l'évolution des positions en temps réel ➤ Connaître la page ayant le meilleur potentiel pour un mot-clé ➤ Une fenêtre d'aide/explicative ➤ Un historique des modifications

Les référenceurs annoncent utiliser **beaucoup d'outils** et aussi **perdre du temps** pour réaliser le diagnostic d'un site web. Ils sont favorables à l'utilisation d'un système d'aide à la décision, néanmoins ils désirent que celui-ci propose des **résultats fiables et pertinents**.

Suite à ce premier questionnaire, nous avons demandé aux référenciers de réaliser le diagnostic de deux sites web différents comportant 15 pages chacun. Pour le premier site, ils devaient réaliser un diagnostic en utilisant leur ancienne méthode (manuelle). Pour réaliser le second diagnostic, ils pouvaient utiliser notre système d'aide à la décision. Nous avons donc mesuré le temps qu'ils passaient pour réaliser le diagnostic « avec » puis « sans » le système d'aide à la décision. Enfin nous leur avons demandé d'attribuer une note entre 1 et 10 (10 étant la meilleure note) pour juger de la pertinence des propositions du système. Les résultats sont présentés dans le tableau 10 ci-dessous :

Tableau 10. Résultat des tests de diagnostic SEO avec et sans l'outil

Référenceur	Temps de diagnostic SANS le système	Temps de diagnostic AVEC le système	Pertinence de l'analyse SANS le système	Pertinence de l'analyse AVEC le système
A	7 min 35	38 secondes	10/10	9/10
B	6 min 35	25 secondes	10/10	9,5/10
C	8 min 14	27 secondes	10/10	8,5/10

Nous remarquons que le temps pour réaliser le diagnostic a **diminué** pour passer d'environ 6-8 minutes à moins d'une minute. Par ailleurs la **pertinence des résultats** obtenus avec l'outil est jugée comme étant bonne, avec des notes de 8,5 à 9,5 sur 10.

Suite à ce test en condition réelle, nous avons distribué notre deuxième questionnaire pour savoir ce que les référenciers ont pensé de ce système. Les réponses sont exposées dans le tableau 11 ci-dessous :

Tableau 11. Deuxième questionnaire

Questions	Réponses des trois référenciers
E. Pensez-vous que ce système d'aide à la décision peut vous aider dans votre travail ?	➤ Oui
F. Comment ce système d'aide à la décision peut-il vous aider ?	<ul style="list-style-type: none"> ➤ Gain de temps ➤ Amélioration de la collaboration avec le client ➤ Diagnostic rapide et clair du travail à effectuer ➤ Permet d'agir sur un(des) critère(s) précis

G. Est-ce que le diagnostic proposé par l'outil vous semble cohérent ?	➤ Oui
H. Comment améliorer ce système d'aide à la décision ?	<ul style="list-style-type: none"> ➤ Améliorer le côté visuel «gamification», notamment pour l'interface client ➤ Proposer une évaluation de la qualité des backlinks (autorité des domaines référents), étude des ancrés des backlinks en rapport avec le mot clé étudié et des ancrés (en maillage interne) ➤ Apporter une idée du volume de recherche/concurrence des mots-clés
I. Pensez-vous que l'outil puisse améliorer la collaboration avec un client ?	<ul style="list-style-type: none"> ➤ Oui, grâce au back-office à disposition du client. Il peut vérifier par lui-même les résultats ➤ Oui, cela permet au client d'avoir un aperçu de son référencement, ce qui est pour lui souvent abstrait ➤ Oui, grâce à l'approche chiffrée et qualitative assurant un suivi pertinent des positions.
J. Seriez-vous intéressé pour utiliser l'outil au quotidien ?	➤ Oui

Les réponses des référenceurs sont positives concernant l'utilisation de cet outil. Ils apprécient le **gain de temps**, la **pertinence des résultats** proposés et l'**aspect collaboratif**. Les points d'amélioration évoqués sont l'**amélioration de l'interface**, la **prise en compte de critères supplémentaires** comme « la qualité des backlinks » et l'**intégration de données supplémentaires** comme le « volume de recherche et la concurrence » d'un mot-clé. Cela peut effectivement améliorer la collaboration avec le client car le référenceur peut appuyer le choix d'un bon mot-clé (volume de recherche élevé et faible concurrence) ou d'un mauvais mot-clé (volume de recherche faible et forte concurrence).

Notre système d'aide à la décision offre une **alternative cohérente** et un **gain de temps** non négligeable au référenceur pour la réalisation du diagnostic d'un site web. Le référenceur pourra désormais proposer des actions correctives au client **en temps réel**. Suite à l'implémentation des actions correctives, le client sera capable de son côté de suivre l'évolution du classement de son site, ce qui lui confère **davantage d'autonomie**.

Le bilan de cet essai en conditions réelles auprès de référenceurs est très positif. Il n'est néanmoins qu'indicatif, car il porte sur un effectif très réduit de référenceurs. Il conviendrait d'élargir ce test sur un panel plus important de personnes pour espérer avoir un résultat significatif.

6. Discussion et conclusion

Dans ce chapitre nous avons pu mettre à l'épreuve notre modèle d'agents présenté au chapitre 4 pour concevoir deux systèmes. Le premier système permet de simuler le classement de sites web afin de **mieux comprendre le fonctionnement du processus de référencement**. Le deuxième système permet au référenceur de gagner du temps et de l'efficacité lors de la phase de « proposition des actions » et ainsi de **mieux maîtriser le processus de référencement**. Ce deuxième système est construit à partir de règles de décision floues. Ces règles ont été définies à partir des résultats issus d'une série d'expérimentations conduites sur les critères *KWrText* (nombre de répétitions du mot-clé dans le texte), *KWrTitle* (nombre de répétitions du mot-clé dans la balise <title>) et *KWrH1* (nombre de répétitions du mot-clé dans la balise <h1>) reconnus pour leur impact significatif sur le classement [22], [23]. Ce système d'aide à la décision a ensuite été testé par les trois référenceurs de notre entreprise pilote qui ont pu valider son utilité et son efficacité.

Depuis leur apparition dans les années 1980 [74], les systèmes d'aide à la décision servent de réel support aux activités nécessitant le recours à un jugement humain. Le référencement est typiquement un métier où **la prise de décision est complexe**, car elle cumule des résultats incertains et un nombre croissant de critères à optimiser [85]. Jankowski et al. [84] ont déjà développé un système d'aide à la décision pour le référencement, mais leur approche demeure différente de la nôtre. En effet, ils ont mené des expérimentations sur des sites web préexistants alors que nous avons volontairement choisi de créer nos propres sites afin d'avoir un meilleur contrôle de tous les paramètres. En effet, sur des sites préexistants, il est difficile, voire impossible, d'avoir des valeurs identiques pour les critères non-testés.

Les expérimentations que nous avons conduites sur des critères reconnus pour leur fort impact sur le classement nous ont permis de définir des règles de décision floues. En effet, le référenceur travaille dans un environnement dynamique ; dès lors le système doit être capable de s'adapter aux changements de manière à rester efficace dans le temps. La logique floue permet d'appréhender les problèmes d'incertitude [86], [87], cela nous a permis de définir des règles de décision suffisamment souples pour progresser en fonction des évolutions des algorithmes de classement. Une fois que le système d'aide à la décision est alimenté par ce type de règles, il est capable de proposer un diagnostic rapide et pertinent au référenceur. Celui-ci peut ensuite suggérer à son client une liste d'actions prioritaires à mettre en place pour améliorer le classement de son site.

En conséquence, nous avons pu répondre à notre troisième hypothèse de recherche, à savoir :

Hypothèse n°3 : « *Le processus de référencement est mieux maîtrisable grâce à l'utilisation d'un système d'aide à la décision* »

En plus du gain de temps obtenu pour réaliser ce diagnostic, le référenceur peut désormais impliquer le client dans le suivi de son classement, ce qui confère à ce dernier davantage d'autonomie. La collaboration entre les deux acteurs n'en sera que meilleure puisqu'ils travaillent sur un référentiel commun [54].

Conclusion générale

1. *Le référencement web est un processus complexe et structurable : apports et limites..* 126
2. *Le processus de référencement web est modélisable : apports et limites* 127
3. *Une meilleure maîtrise du référencement web est possible grâce à un système d'aide à la décision : apports et limites* 128

L'objectif de notre travail de thèse était double : d'une part, permettre une **meilleure compréhension** du processus de référencement ; d'autre part, procurer aux référenceurs un système capable de servir de support à leur activité pour répondre de manière adéquate aux besoins du client et ainsi permettre une **meilleure maîtrise** de ce processus de référencement.

Après avoir réalisé un état de l'art sur le référencement, nous avons constaté que cette activité est étroitement liée à l'évolution des algorithmes de classement des moteurs de recherche. En effet, pour proposer des résultats toujours plus pertinents à leurs utilisateurs afin de rester compétitifs, les moteurs de recherche doivent continuellement améliorer leurs méthodes de classement. Ce constat révèle une **structure de travail dynamique et en perpétuelle évolution**. C'est dans ce contexte particulier qu'est née l'activité de référencement. Réservée au départ à une poignée d'autodidactes, elle s'est peu à peu professionnalisée avec la création de programmes de formation dispensés par des structures publiques ou privées, au point de devenir aujourd'hui un métier pérenne [4] à l'image des activités conduites au sein de notre entreprise pilote où nos travaux ont été conduits. En une vingtaine d'années, les référenceurs ont pu bâtir les fondements de la profession grâce à leur grande **capacité d'adaptation** et leur parfaite **connaissance des moteurs de recherche**.

Néanmoins, la pratique de cette activité demeure en réalité très artisanale, car **l'absence de cadre** et le **recours à l'expérience** sont légion. De plus, les résultats sont relativement **incertains** ce qui oblige les référenceurs à prendre du recul. Ils doivent aussi constamment adapter leur niveau de discours selon les clients en privilégiant la qualité des interactions, car celles-ci sont très **hétérogènes**. Les référenceurs doivent aussi analyser un **volume important de critères** ce

qui les oblige à **se tenir continuellement informés**. Par ailleurs, comme nous avons pu le constater dans notre entreprise pilote, les **impératifs de temps et de coût** à respecter ne leur permettent pas toujours d'optimiser l'ensemble des critères d'un site web. Enfin les outils à leur disposition sont nombreux, voire trop nombreux, ce qui les oblige à rester méthodiques pour ne pas se perdre dans ce **flot important d'informations**. L'ensemble de ces difficultés nous a amené à considérer le référencement comme un **problème complexe**.

Dans les sections suivantes, nous présentons les apports et limites de chacune de nos contributions, puis nous terminerons par la présentation de nos perspectives de travail.

1. **Le référencement web est un processus complexe et structurable : apports et limites**

Le référenceur est un acteur humain qui transforme les éléments d'entrée (pages web mal classées) en éléments de sortie (pages web bien classées) grâce à l'action de référencement. Pour ce faire, le référenceur doit trouver des solutions adéquates pour satisfaire les besoins et les attentes de son client en tenant compte des contraintes de l'environnement (évolution des algorithmes de classement, technologie utilisée sur le site web, etc.) Malgré son caractère artisanal [4], nous pensons que l'activité de référencement peut être structurée. Ce constat nous a amené à considérer le référencement comme un problème complexe nécessitant la définition d'un processus [39], et plus particulièrement d'un processus de conception [40], [46]. Ceci est notre première contribution majeure. En effet, jusqu'à maintenant la littérature scientifique déjà limitée sur le sujet se cantonnait à identifier les bonnes ou les mauvaises pratiques pour optimiser le classement d'un site web, limitant l'activité de référencement à une simple « combinaison de techniques » ainsi que nous avons pu le montrer [37], [38]. Nous pensons au contraire que le référencement peut être comparé à une activité d'ingénierie. Dès lors, l'utilisation d'un méta-modèle d'ingénierie pour identifier les étapes du processus de référencement est possible. Grâce à cette approche, nous avons pu mettre en évidence cinq étapes clés dans le processus de référencement, à savoir : (1) l'identification du besoin, (2) l'étude de faisabilité, (3) la

proposition des actions, (4) la mise en place des actions et (5) la mesure des résultats, assez comparables à celles décrites dans les processus de conception [41], [42]. En conséquence, nous avons pu répondre favorablement à notre première hypothèse, à savoir :

Hypothèse n°1 : « *Le processus de référencement est un processus complexe et structurable* »

La limite de notre approche serait de considérer le référencement comme un processus purement séquentiel, avec des phases d'études qui se suivent chronologiquement, souvent qualifié de processus statique, dont l'évolution serait limitée. En effet, le référencement est une activité en perpétuel changement, qui nécessite de remettre en question ses connaissances, l'ordre de ses phases d'étude, mais aussi ses méthodes de travail. Dès lors, nous privilégions le recours à un processus qualifié de dynamique.

2. Le processus de référencement web est modélisable : apports et limites

Nous avons identifié plusieurs entités distribuées au sein de processus de référencement : les *besoins*, les *fonctions*, les *contraintes* et les *solutions*. Nous savons que le paradigme agent est particulièrement adapté pour modéliser des systèmes distribués [56]. En effet, les caractéristiques des agents (distribution, autonomie, flexibilité, coopération) nous ont permis de répondre de manière adéquate à nos besoins en termes de modélisation et de résolution de problème distribué pour le référencement. Par ailleurs, cette approche nous a permis d'identifier des groupes d'agents autonomes comme les *Pages* ou les *Critères* qui jouent un rôle majeur dans la prise de décision. L'apport de notre travail pour la communauté scientifique a été d'utiliser le paradigme agent pour faciliter et modéliser l'activité de référencement. Ceci est notre deuxième contribution majeure. Jusqu'à maintenant les systèmes multi-agents étaient utilisés dans beaucoup de domaines comme les transports [88], la conception [63] et la finance [89]. À notre connaissance les systèmes multi-agents n'ont été utilisés qu'une seule fois pour améliorer la maîtrise du référencement, [84] mais l'approche qui a été conduite est différente de la nôtre puisque les expérimentations ont été menées sur des sites web

préexistants. En conséquence cela nous a permis de répondre positivement à notre deuxième hypothèse de travail, à savoir :

Hypothèse n°2 : « *Le processus de référencement est modélisable* »

Les systèmes multi-agents ne sont sans doute pas le seul outil permettant la modélisation du processus de référencement. Comme nous l'avons vu, les réseaux de Petri, les réseaux de neurones artificiels et les algorithmes génétiques peuvent aussi être adaptés à ce genre de démarche [65], [66]. Mais de notre point de vue, les systèmes multi-agents proposent le **meilleur compromis** pour répondre à l'ensemble de nos besoins en termes de modélisation.

3. Une meilleure maîtrise du référencement web est possible grâce à un système d'aide à la décision : apports et limites

Nous avons mis à contribution notre modèle d'agents pour concevoir un système d'aide à la décision servant de support aux référenceurs. L'apport de notre travail à la communauté scientifique a été de concevoir un système d'aide à la décision pour le référencement utilisant des règles de décision floues définies à partir d'une série d'expérimentations. À notre connaissance aucun système d'aide à la décision utilisant des règles floues dédié au référencement n'a été développé à ce jour. Nous savons que d'autres outils ont déjà été développés pour aider le référenceur, ainsi Su et al. [90] proposent un système permettant de prédire le classement d'un site web, Zhu et al. [91] présentent un système d'édition de contenu basé sur la collecte de mots-clés pertinents, et Di et al. [92] proposent un système de recommandation de mots-clés et de contrôle de backlinks. Comme les outils que nous avons présentés dans le chapitre 1, la plupart d'entre eux sont conçus pour un domaine d'action précis et contraignent souvent le référenceur à devoir utiliser plusieurs outils pour exercer son activité. De plus, ils n'intègrent pas le rôle du client dans l'activité de référencement. Dans notre système, nous procurons au référenceur un diagnostic instantané des actions correctives à mettre en place pour améliorer le classement d'un site web. Ce support, en plus de lui offrir un gain de temps non négligeable pour la phase de « propositions des actions », lui permet d'avoir une meilleure maîtrise du processus de référencement. En conséquence, ces résultats nous ont permis de

répondre positivement à notre troisième hypothèse de travail, à savoir :

Hypothèse n°3 : « *Le processus de référencement est mieux maîtrisable grâce à l'utilisation d'un système d'aide à la décision* »

Les règles de décision définies à partir d'expérimentations sont tributaires de la dimension temporelle. C'est-à-dire que pour garder des règles cohérentes avec la réalité du terrain, de nouvelles expérimentations doivent être effectuées régulièrement.

En répondant favorablement à ces trois hypothèses de travail, nous avons été capables de répondre à notre problématique de recherche qui était :

Problématique : « *Comment concevoir le processus de référencement pour améliorer sa maîtrise par le couple référenceur/client ?* »

Perspectives envisagées

1. *Vers des systèmes intelligents et coopératifs pour assister l'ingénierie du référencement*
130
2. *Vers une approche multi-critères versus critères forts.....* 133
3. *Vers l'intégration des contraintes dynamiques de coût et de temps passé* 134

Les recherches que nous avons conduites dans le cadre de cette thèse nous ont permis d'envisager trois perspectives de travail. Nous présentons chacune d'elles ci-dessous.

1. **Vers des systèmes intelligents et coopératifs pour assister l'ingénierie du référencement**

Notre recherche a contribué à transformer « l'art » du référencement en ingénierie du référencement. Dans sa version actuelle, notre système d'aide à la décision exige une maintenance régulière, car il doit évoluer en accord avec les changements d'algorithmes des moteurs de recherche. Ainsi, pour que le diagnostic proposé reste pertinent, nous devons constamment conduire de nouvelles expérimentations pour maintenir le système à jour. Sur la base des expérimentations que nous avons menées pour tester les critères du référencement, nous avons envisagé maintes fois **d'automatiser la procédure** ; c'est-à-dire que la création des sites web expérimentaux puisse se faire de manière automatisée. En effet, ces expérimentations sont très chronophages, surtout quand les critères doivent être combinés. Par ailleurs, comme nous l'avons constaté lors de nos expérimentations, certains critères testés (ex : *KWrTitle* et *KWrHI*) exigent davantage de temps pour se stabiliser (minimum 3 semaines) et pour pouvoir engendrer des résultats exploitables (1 mois supplémentaire), ce qui rallonge significativement la durée des expérimentations.

Nous pensons que l'achat anticipé de noms de domaine et la création automatisée de pages web permettraient de gagner du temps. Mais il faut veiller à toujours utiliser du texte généré aléatoirement pour se prémunir contre les pénalités liées à la duplication de contenu et initialiser les valeurs non-testées de manière identique pour maîtriser l'ensemble des paramètres de l'expérimentation.

Voici le protocole que nous pensons mettre en place pour tester l'automatisation des expérimentations :

1. Stockage de contenus texte générés aléatoirement dans une base de données
2. Achat automatisé de noms de domaine / hébergements³⁵
3. Création automatique de pages web pour chaque nom de domaine grâce à des scripts PHP (les scripts insèrent les mots-clés dans les pages aux emplacements prédéfinis)

Si ce protocole est bien adapté pour automatiser les expérimentations sur davantage de critères *on-page*, il sera néanmoins plus difficile de le mettre en œuvre pour tester des critères *off-page*. En effet, pour cela il faudrait maîtriser des paramètres situés en dehors de notre champ expérimental (ex : nombre de Like sur Facebook).

Finalement, en automatisant la partie expérimentale, les résultats obtenus seraient mis à jour régulièrement ce qui permettrait de garantir la pertinence des règles de décision tout en les prémunissant contre l'obsolescence. Mais pour réduire davantage la durée de maintenance du système, l'automatisation des expérimentations n'est pas suffisante. En effet, les règles de décision sont créées et modifiées manuellement. Dès lors, automatiser les expérimentations et ainsi augmenter leur volume, n'auraient que peu d'intérêt si la base de règles doit constamment être modifiée manuellement.

Nous avons donc envisagé de rendre notre système d'aide à la décision **proactif**. L'idée est que le système puisse définir ou modifier lui-même des règles de décision à partir des résultats issus des expérimentations. Pour ce faire, nous avons imaginé un protocole en trois étapes :

1. Automatiser l'expérimentation n (voir protocole précédent)

³⁵ Certains hébergeurs proposent des scripts permettant d'automatiser la procédure d'achat, à l'image de l'hébergeur OVH : https://www.ovh.com/fr/domaines/service_soapi.xml

2. Analyser les résultats de l'expérimentation n (l'analyse du classement peut se faire via l'outil MyPoseo ; les positions peuvent être stockées dans une nouvelle base de données « Résultats expérimentations »).
3. Si les résultats obtenus sont différents de l'expérimentation $n-1$ alors le système doit « mettre à jour la règle existante » ou « créer une nouvelle règle de décision »

Seule une intervention manuelle serait nécessaire pour la définition des ensembles flous. Néanmoins, nous pensons qu'il est possible de définir à l'avance des ensembles flous en fonction d'intervalles de classement prédéfinis (ex : *classement excellent*, entre 1^{ière} et 3^e position ; *classement moyen* : entre 4^e et 10^e position ; *mauvais classement* : à partir de la 11^e soit classiquement la deuxième page). Cela permettrait de s'abstenir de toute intervention manuelle pour la génération des règles de décision.

Notons par ailleurs que cette perspective visant à automatiser les expérimentations et à rendre le système proactif nécessite de lever et de reconcevoir le lien entre les différents scripts (hébergeur, création de pages, analyse des résultats, création/modification des règles de décision, création/modification des ensembles flous). Cela exige également une transformation complète du *Moteur SEO* pour qu'il puisse fonctionner de manière autonome.

Si une approche multi-critères plus fine est à mettre en place, une autre perspective nous semble indispensable pour améliorer les performances et la pertinence de l'outil : la conception d'un véritable Système d'Aide à la Décision Coopératif, exploitant les capacités coopératives des agents le composant. En réaction continue aux évolutions du référencement, le système pourrait alors proposer des décisions pondérées au couple client/référenceur, ce qui leur permettrait de superviser et d'adapter de manière itérative et incrémentale le référencement. L'outil, rendu ainsi plus intelligent, jouerait pleinement son rôle d'assistant et de médiateur de référencement. Une perspective à plus long terme serait de construire un système intelligent d'apprentissage pour assister l'ingénierie du référencement. Ces deux dernières évolutions permettraient à notre système d'appartenir à la catégorie des SAD coopératifs.

2. Vers une approche multi-critères versus critères forts

Dans notre étude nous nous sommes volontairement limités à mener des expérimentations sur trois critères de type *on-page*. En effet, ces trois critères sont connus pour leur fort impact sur le classement [22], [23], ils permettent d'obtenir des résultats rapides et fiables, et ils sont couramment utilisés par les référenceurs et en particulier par les référenceurs de notre entreprise pilote. Mais soulignons que notre système a été conçu pour fonctionner avec **un panel beaucoup plus large de critères** afin d'ajouter de nouvelles règles et ainsi perfectionner encore la pertinence du diagnostic proposé. Si les critères testés jusqu'à maintenant étaient des critères *on-page*, nous souhaitons également tester des critères *off-page*. Dans un premier temps nous aimerions conduire des expérimentations sur les 15 critères identifiés lors de la construction du système de classement expérimental, car leur impact sur le classement est significatif [7], [8], [19], [21]–[23], [28].

Cette deuxième perspective pour une approche multi-critères plus étendue est déjà engagée, les expérimentations en cours anticipent d'ailleurs la mise en place du protocole d'automatisation. Nous attendons ainsi de ces futures expérimentations l'ajout de nouveaux critères, une meilleure définition de nos règles de décision et l'amélioration de la pertinence du diagnostic proposé par le système. Une perspective à moyen terme serait aussi de développer des modèles qui permettent d'extraire les critères (variables) de contrôle, au sens thomien³⁶, du processus de référencement. L'établissement d'une classification plus fine de ces critères ainsi qu'une meilleure évaluation des critères prépondérants dans le référencement fait aussi partie de cette perspective.

³⁶ René THOM : Mathématicien et philosophe français, fondateur de la théorie des catastrophes. Il a reçu la médaille Fields en 1958.

3. Vers l'intégration des contraintes dynamiques de coût et de temps passé

Lors des essais réalisés auprès des trois référenceurs de notre entreprise pilote, ceux-ci nous ont signalé que pendant le diagnostic d'un site web, leur expérience leur permettait (contrairement à notre système) de réaliser une évaluation du tarif de la future prestation. Notre solution aujourd'hui ne donne aucune indication sur les coûts estimés par la mise en place des actions correctives sur un site web.

Dans notre entreprise pilote, le tarif d'une prestation de référencement dépend de plusieurs facteurs : le nombre de mots-clés à référencer, la concurrence sur les mots-clés choisis, le nombre de critères à corriger et le temps passé pour corriger chaque critère. Notre troisième perspective de travail serait d'intégrer ces différents facteurs dans notre système d'aide à la décision pour qu'il soit capable de proposer une estimation de chiffrage au référenceur.

Ainsi pour augmenter la maîtrise du processus de référencement, non seulement sur les solutions techniques définies, nous proposons d'intégrer la dimension économique à notre système, ce qui nécessite l'élaboration de modèles mathématiques spécifiques permettant une meilleure gestion des coûts et du temps. Enfin, une validation du système à plus grande échelle nous permettra d'identifier davantage de contraintes que ne manqueront pas de nous formuler les référenceurs participants au test.

Bibliographie

- [1] D. Hawking, « Web search engines. Part 1 », *Computer*, vol. 39, n° 6, p. 86-88, juin 2006.
- [2] D. Hawking, « Web Search Engines: Part 2 », *Computer*, vol. 39, n° 8, p. 88-90, août 2006.
- [3] N. Yalçın et U. Köse, « What is search engine optimization: SEO? », *Procedia - Social and Behavioral Sciences*, vol. 9, n° 0, p. 487-493, 2010.
- [4] J.-C. Domenget, « Le métier de référenceur est-il pérenne et légitime en communication? », *Les cahiers du Résiproc*, n° 2, 2014.
- [5] J. Caelen, *Le consommateur au coeur de l'innovation*. CNRS, 2013.
- [6] S. Sagot, A.-J. Fougères, et E. Ostrosi, « A multi-agent approach for building a fuzzy decision support system to assist the SEO process », in *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016, vol. 2016.
- [7] O. Andrieu, L. Bourelly, et Grifil, *Réussir son référencement web: stratégie et techniques SEO*. Eyrolles, 2016.
- [8] M. Chartier, *Le guide du référencement web*. First interactive, 2013.
- [9] J.-P. Lardy, « Méthodes de tri des résultats des moteurs de recherche », *La lettre de l'URFIST*, 2000.
- [10] S. Brin et L. Page, « The anatomy of a large-scale hypertextual Web search engine », *Computer networks and ISDN systems*, vol. 30, n° 1, p. 107-117, 1998.
- [11] J. M. Kleinberg, « Authoritative sources in a hyperlinked environment », *Journal of the ACM (JACM)*, vol. 46, n° 5, p. 604-632, 1999.
- [12] Z. Gyöngyi, H. Garcia-Molina, et J. Pedersen, « Combating web spam with trustrank », in *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*, 2004, p. 576-587.
- [13] R. Guha, *Search result ranking based on trust*. Google Patents, 2009.
- [14] M. Curtiss, K. Bharat, et M. Schmitt, *Systems and methods for improving the ranking of news articles*. Google Patents, 2009.
- [15] Y. Liu *et al.*, « BrowseRank: letting web users vote for page importance », in *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, 2008, p. 451-458.
- [16] M. Koster, « ALIWEB-Archie-like indexing in the Web », *Computer Networks and ISDN Systems*, vol. 27, n° 2, p. 175-182, 1994.
- [17] Eyetools, Enquiro, et Did-it, « Eyetools, Enquiro, and Did-it uncover Search's Golden Triangle. », 2005. [En ligne]. Disponible sur: <http://www.prweb.com/releases/2005/03/prweb213516.htm>.

- [18] A. Ghose et S. Yang, « An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets », *Management Science*, vol. 55, n° 10, p. 1605-1622, oct. 2009.
- [19] S. O’Neill et K. Curran, « The Core Aspects of Search Engine Optimisation Necessary to Move up the Ranking », *International Journal of Ambient Computing and Intelligence (IJACI)*, vol. 3, n° 4, p. 62-70, 2011.
- [20] K. Curran, « Tips for Achieving High Positioning in the Results Pages of the Major Search Engines », *Information Technology Journal*, vol. 3, n° 2, p. 202-205, 2004.
- [21] M. P. Evans, « Analysing Google rankings through search engine optimization data », *Internet Research*, vol. 17, n° 1, p. 21-37, 2007.
- [22] J. Zhang et A. Dimitroff, « The Impact of Webpage Content Characteristics on Webpage Visibility in Search Engine Results (Part I) », *Information Processing & Management*, vol. 41, n° 3, p. 665-690, 2005.
- [23] J. Zhang et A. Dimitroff, « The impact of metadata implementation on webpage visibility in search engine results (Part II) », *Information Processing & Management*, vol. 41, n° 3, p. 691-715, 2005.
- [24] F. Wang, Y. Li, et Y. Zhang, « An empirical study on the search engine optimization technique and its outcomes », in *Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), 2011 2nd International Conference on*, 2011, p. 2767-2770.
- [25] C. Zhu et G. Wu, « Research and Analysis of Search Engine Optimization Factors Based on Reverse Engineering », 2011, p. 225-228.
- [26] P. Patil Swati, B. V. Pawar, et S. Patil Ajay, « Search Engine Optimization: A Study ».
- [27] K. Dye, « Website abuse for search engine optimisation », *Network Security*, vol. 2008, n° 3, p. 4-6, 2008.
- [28] R. A. Malaga, « Search engine optimization—black and white hat approaches », *Advances in Computers*, vol. 78, p. 1-39, 2010.
- [29] G. Sire, « Référencement et référencement. Cachez ces pratiques que je ne saurais voir », *Sur le journalisme About journalism Sobre jornalismo*, vol. 3, n° 1, p. 70-83, 2014.
- [30] A. Somani et U. Suman, « Counter measures against evolving search engine spamming techniques », in *Electronics Computer Technology (ICECT), 2011 3rd International Conference on*, 2011, vol. 6, p. 214-217.
- [31] A. Ntoulas, M. Najork, M. Manasse, et D. Fetterly, « Detecting spam web pages through content analysis », in *Proceedings of the 15th international conference on World Wide Web*, 2006, p. 83-92.
- [32] P. Swati, B. V. Pawar, et S. Patil Ajay, « Search Engine Optimization: A Study », 2012.
- [33] R. A. Malaga, « Worst Practices in Search Engine Optimization », *Commun. ACM*, vol. 51, n° 12, p. 147-150, 2008.

- [34] D. Y. Wang, S. Savage, et G. M. Voelker, « Cloak and dagger: dynamics of web search cloaking », in *Proceedings of the 18th ACM conference on Computer and communications security*, 2011, p. 477–490.
- [35] F. Raiber, K. Collins-Thompson, et O. Kurland, « Shame to be sham: addressing content-based grey hat search engine optimization », in *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, 2013, p. 1013–1016.
- [36] M. Lewkowicz et M. Zacklad, « MEMO-net, un collecticiel utilisant la méthode de résolution de problème DIPA pour la capitalisation et la gestion des connaissances dans les projets de conception », *IC*, vol. 99, p. 119–128, 1999.
- [37] S. Sagot, A.-J. Fougères, E. Ostrosi, et P. Lacom, « Search engine optimization: From analysis based on an engineering meta-model towards integrative approaches », in *Proceedings of International Conference on Information Society (i-Society)*, 2014, p. 274–281.
- [38] R. A. Malaga, « The value of search engine optimization: An action research project at a new e-commerce site », *Journal of Electronic Commerce in Organizations (JECO)*, vol. 5, n° 3, p. 68–82, 2007.
- [39] AFNOR, « NF EN ISO 9000:2015 : Systèmes de management de la qualité — Principes essentiels et vocabulaire », Norme, oct. 2015.
- [40] AFNOR, « NF EN ISO 9001:2015 : Systèmes de management de la qualité — Exigences », Norme, oct. 2015.
- [41] V. Hubka et W. E. Eder, *Design Science*. London: Springer London, 1996.
- [42] N. Cross, J. Naughton, et D. Walker, « Design method and scientific method », *Design studies*, vol. 2, n° 4, p. 195–201, 1981.
- [43] N. P. Suh, « Axiomatic design theory for systems », *Research in engineering design*, vol. 10, n° 4, p. 189–209, 1998.
- [44] N. P. Suh, *The principles of design*, vol. 990. Oxford University Press New York, 1990.
- [45] E. Ostrosi, A.-J. Fougères, M. Ferney, et D. Klein, « A fuzzy configuration multi-agent approach for product family modelling in conceptual design », *Journal of Intelligent Manufacturing*, vol. 23, n° 6, p. 2565–2586, 2012.
- [46] AFNOR, « FD X50-127 : Outils de management - Maîtrise du processus de conception et développement », Norme, avr. 2002.
- [47] E. Ostrosi, D. Choulier, et M. Kurth, « A Meta-Model for Engineering Analysis in Product Design », in *Concurrent Engineering Approaches for Sustainable Product Development in a Multi-Disciplinary Environment*, J. Stjepandić, G. Rock, et C. Bil, Éd. Springer London, 2013, p. 545–556.
- [48] E. Ostrosi, A.-J. Fougères, et M. Ferney, « Fuzzy agents for product configuration in collaborative and distributed design process », *Applied Soft Computing*, vol. 12, n° 8, p. 2091–2105, 2012.
- [49] E. Ostrosi, L. Haxhijaj, et S. Fukuda, « Fuzzy modelling of consensus during design conflict resolution », *Research in Engineering Design*, vol. 23, n° 1, p. 53–70, janv. 2012.

- [50] D. Brissaud et O. Garro, « An approach to concurrent engineering using distributed design methodology », *Concurrent Engineering*, vol. 4, n° 3, p. 303–311, 1996.
- [51] G. Prudhomme, J.-F. Boujut, et F. Pourroy, « Activité de conception et instrumentation de la dynamique des connaissances locales », in *Ingénierie des connaissances*, 2005, p. 413–436.
- [52] E. Blanco, « L'émergence du produit dans la conception distribuée vers de nouveaux modes de rationalisation dans la conception de systèmes mécaniques », Thèse, Université Joseph-Fourier, Grenoble, 1998.
- [53] M. Zacklad, « La théorie des Transactions Intellectuelles: une approche gestionnaire et cognitive pour le traitement du COS », *Intellectica*, vol. 1, n° 30, p. 195–222, 2000.
- [54] J. Hoc, « Coopération humaine et systèmes coopératifs », *Ingénierie cognitive. IHM et cognition*, p. 139–187, 2003.
- [55] M. Zacklad, M. Lewkowicz, J.-F. Boujut, F. Darses, et F. Détienne, « Formes et gestion des annotations numériques collectives en ingénierie collaborative », *Dieng, R., éditeur: IC*, vol. 2003, p. 207–224, 2003.
- [56] G. Weiss, Éd., *Multiagent systems: a modern approach to distributed artificial intelligence*. Cambridge, Mass: MIT Press, 1999.
- [57] E. H. Durfee et J. S. Rosenschein, « Distributed problem solving and multi-agent systems: Comparisons and examples », *Ann Arbor*, vol. 1001, n° 48109, p. 29, 1994.
- [58] K. Slimani, D. Pallez, L. Médini, et P. Ghodous, « Multi-agent system architecture for concurrent design and manufacturing », in *11th ISPE International Conference on Concurrent Engineering*, 2004, p. 209–215.
- [59] R. D. Beer, « A dynamical systems perspective on agent-environment interaction », *Computational Theories of Interaction and Agency*, vol. 72, p. 173–215, 1996.
- [60] M. Wooldridge, N. R. Jennings, et others, « Intelligent agents: Theory and practice », *Knowledge engineering review*, vol. 10, n° 2, p. 115–152, 1995.
- [61] M. Wooldridge, « Agent-based software engineering », in *Software Engineering. IEEE Proceedings*, 1997, vol. 144, p. 26–37.
- [62] N. R. Jennings, « On agent-based software engineering », *Artificial intelligence*, vol. 117, n° 2, p. 277–296, 2000.
- [63] L. Monostori, J. Váncza, et S. R. Kumara, « Agent-based systems for manufacturing », *CIRP Annals-Manufacturing Technology*, vol. 55, n° 2, p. 697–720, 2006.
- [64] J. Ferber, *Multi-agent systems: an introduction to distributed artificial intelligence*, vol. 1. Addison-Wesley Reading, 1999.
- [65] Q. Zhu et A. T. Azar, *Complex system modelling and control through intelligent soft computations*. Springer, 2015.
- [66] E.-G. Talbi, *Metaheuristics: from design to implementation*, vol. 74. John Wiley & Sons, 2009.
- [67] C. M. Macal et M. J. North, « Tutorial on agent-based modelling and simulation », *Journal of simulation*, vol. 4, n° 3, p. 151–162, 2010.

- [68] N. Boccarda, *Modeling complex systems*. Springer Science & Business Media, 2010.
- [69] J.-P. Müller, « Des systèmes autonomes aux systèmes multi-agents: Interaction, émergence et systèmes complexes », Université Libre de Bruxelles, 2002.
- [70] A.-J. Fougères, « Vers un système de médiation pour les systèmes coopératifs », Université de Technologie de Compiègne, 2010.
- [71] A.-J. Fougères, « Agents to cooperate in distributed design », in *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, 2004, vol. 3, p. 2629–2634.
- [72] A.-J. Fougères, « Modelling and simulation of complex systems: an approach based on multi-level agents », *arXiv preprint arXiv:1201.3880*, 2012.
- [73] A.-J. Fougères, « Modelling and simulation of complex systems: an approach based on multi-level agents », *International Journal of Computer Science Issues*, vol. 8, n° 6, p. 8–17, 2011.
- [74] I. Nizetić, K. Fertalj, et B. Milašinović, « An Overview of Decision Support System Concepts », in *18th International Conference on Information and Intelligent Systems*, 2007, p. 251–256.
- [75] P. Brézillon, J.-C. Pomerol, et I. Saker, « Contextual and contextualized knowledge: An application in subway control », *International Journal of Human-Computer Studies*, vol. 48, n° 3, p. 357–373, 1998.
- [76] H. Omrani, A. Ogor, L. Ion-Boussier, et P. Trigano, « A Web-Based Decision Support System for Impacts Assessment of Urban Mobility: WDSS4IA », in *Advances in Intelligent Web Mastering*, Springer, 2007, p. 272–277.
- [77] H. Omrani, A. Awasthi, L. Ion, et P. Trigano, « A hybrid approach for evaluating environmental impacts for urban transportation mode sharing », *Journal of Decision Systems*, vol. 18, n° 2, p. 185–201, 2009.
- [78] D. Y. Shee et Y.-S. Wang, « Multi-criteria evaluation of the web-based e-learning system: A methodology based on learner satisfaction and its applications », *Computers & Education*, vol. 50, n° 3, p. 894–905, 2008.
- [79] N. Cavus, « The application of a multi-attribute decision-making algorithm to learning management systems evaluation », *British Journal of Educational Technology*, vol. 42, n° 1, p. 19–30, 2011.
- [80] E. Popescu, P. Trigano, et C. Badica, « Towards a unified learning style model in adaptive educational systems », in *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*, 2007, p. 804–808.
- [81] P. Brézillon et J.-C. Pomerol, « User acceptance of interactive systems: Lessons from knowledge-based and decision support systems », *Failure and Lessons Learned in Information Technology Management*, vol. 1, n° 1, p. 67–75, 1997.
- [82] G. Morvan, « Approche multi-agents d'un système d'aide à la décision en environnement dynamique incertain-Application à l'entomologie médico-légale », Artois, 2009.
- [83] V. E. Ospina Becerra, « Vers une méthodologie de conception de systèmes de médiation adaptés au travail coopératif », Troyes, 2007.

- [84] J. Jankowski et K. Grzasko, « Multi-agent decision support system for search engines marketing », *POLISH ASSOCIATION FOR KNOWLEDGE MANAGEMENT, Studies & Proceedings*, vol. 21, 2009.
- [85] S. Sagot, A.-J. Fougères, et E. Ostrosi, « Search Engine Optimization Process: A Concurrent Intelligent Computing Approach », *Proceedings of 22nd ISPE International Conference on Concurrent Engineering (CE)*, p. 603–614, 2015.
- [86] L. A. Zadeh, « Fuzzy sets », *Information and control*, vol. 8, n° 3, p. 338–353, 1965.
- [87] L. A. Zadeh, « Outline of a new approach to the analysis of complex systems and decision processes », *IEEE Transactions on systems, Man, and Cybernetics*, n° 1, p. 28–44, 1973.
- [88] F. Balbo et S. Pinson, « Toward a multi-agent modelling approach for urban public transportation systems », in *International Workshop on Engineering Societies in the Agents World*, 2001, p. 160–174.
- [89] T. Lux et M. Marchesi, « Scaling and criticality in a stochastic multi-agent model of a financial market », *Nature*, vol. 397, n° 6719, p. 498–500, 1999.
- [90] A.-J. Su, Y. C. Hu, A. Kuzmanovic, et C.-K. Koh, « How to improve your Google ranking: Myths and reality », in *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, 2010, vol. 1, p. 50–57.
- [91] X. Zhu et Z. Tan, « SEO Keyword Analysis and Its Application in Website Editing System », in *Wireless Communications, Networking and Mobile Computing (WiCOM), 2012 8th International Conference on*, 2012, p. 1–4.
- [92] W. Di, L. Tian, B. Yan, W. Liyuan, et L. Yanhui, « Study on SEO monitoring system based on keywords & links », in *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on*, 2010, vol. 5, p. 450–453.

Liste de nos publications sur le travail conduit

Communications avec actes dans un congrès international :

S. Sagot, A.-J. Fougères, E. Ostrosi, et P. Lacom, « Search engine optimization: From analysis based on an engineering meta-model towards integrative approaches », in Proceedings of International Conference on Information Society (i-Society), 2014, p. 274–281.

S. Sagot, A.-J. Fougères, et E. Ostrosi, « Search Engine Optimization Process: A Concurrent Intelligent Computing Approach », in Proceedings of 22nd ISPE International Conference on Concurrent Engineering (CE), p. 603–614, 2015.

S. Sagot, A.-J. Fougères, et E. Ostrosi, « A multi-agent approach for building a fuzzy decision support system to assist the SEO process », in IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2016, vol. 2016.

Résumé :

Le développement rapide du World Wide Web depuis les années 1990 a entraîné l'accroissement du nombre de sites web et l'intensification de la concurrence. Parallèlement, les moteurs de recherche ont joué un rôle majeur en facilitant la recherche d'informations sur le web, au point de devenir une source inépuisable de trafic dont les propriétaires de sites web souhaitent profiter. Afin de capter un maximum de trafic qualifié, il est devenu nécessaire de se démarquer pour améliorer la visibilité d'un site web sur les moteurs de recherche. Pour cela les propriétaires de sites web peuvent faire appel à des entreprises pratiquant le référencement web, et plus particulièrement le référencement naturel qui regroupe l'ensemble des techniques d'optimisation permettant d'indexer et de positionner les pages d'un site web au plus haut niveau dans les résultats de recherche. L'activité de référencement web est conduite par le référencier qui, en optimisant un certain nombre de critères au sein du site web, améliore son positionnement. Mais les résultats obtenus sont incertains, car les algorithmes de classement des moteurs de recherche sont en perpétuelle mutation. Dès lors, il est devenu nécessaire pour le référencier d'avoir une meilleure maîtrise du processus de référencement web et d'être aidé dans ses prises de décisions afin de satisfaire ses clients. Nos travaux de recherche ont permis d'apporter trois contributions :

- **Établissement d'un cadre efficace** pour l'activité de référencement à partir de la définition des différentes étapes du processus de référencement web en s'inspirant d'un méta-modèle d'ingénierie ;
- **Modélisation du processus de référencement web** à partir d'un modèle d'agents afin de permettre une meilleure compréhension de son fonctionnement et de contrôler efficacement les différentes interactions ;
- **Conception d'un système d'aide à la décision** à partir de notre modèle d'agents servant de support à l'activité de référencement web.

Ces trois contributions majeures nous ont permis de répondre à notre problématique de recherche qui se proposait de répondre à la question suivante : « *Comment concevoir le processus de référencement web pour améliorer sa maîtrise par le couple référencier/client ?* ».

Mots clés : Référencement web, processus de référencement web, système d'aide à la décision, système multi-agents

Abstract:

The World Wide Web has quickly grown since the 90s, and it has led to an increase in the number of websites and the intensification of the competition. In parallel, search engines have played a major role by facilitating the research of information on the web, to the point of being an infinite source of traffic of which websites' owners would like to take advantage. In order to capture the maximum of qualified traffic, it has become necessary to stand out from the competitors to improve the visibility of a website on search engines. To do that, websites' owners can resort to firms which practice Search Engine Optimization (SEO). The latter gathers all the optimization techniques which make it possible to improve the ranking of web pages at the highest rank in the search results. The activity of search engine optimization is done by a SEO practitioner who, by optimizing some technical criteria, improves the ranking of the website. But the obtained results are uncertain, because of the ever-changing ranking algorithms of search engines. Consequently, it has become necessary for the SEO practitioner to have a better mastery of the SEO process and to be helped in its decision makings in order to satisfy its clients. Our research works have allowed us to make three contributions:

- **Creation of an efficient structure** for the SEO activity from the definition of the different steps of the SEO process and the use of an engineering meta-model;
- **Modeling of the SEO process** by using a multi-agent system in order to enable a better understanding of its functioning and to efficiently master its different interactions;
- **Design of a decision support system** thanks to our agent-based model to support the SEO activity.

These three main contributions enabled us to answer our research question which was: "How to design the search engine optimization process in order to improve its mastery by the SEO practitioner and the client?"

Keywords: SEO, Search Engine Optimization, Search Engine Optimization process, Decision Support System, Multi-Agent System

The logo for SPIM (École doctorale SPIM) features the letters 'S', 'P', 'I', and 'M' in a large, white, sans-serif font. The 'S' is stylized with a blue horizontal bar behind it.