الجمـــهوريـــة الجـــزائريـــة الديـمقـراطيـــــــة الشـــعبيـــــة
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعـلـيـم الـعـالـــي و الـبـحـث الـعـلـمي
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
جـامـعـة عمّـــار ثـلـيـــجـــي بالأغـــواط
UNIVERSITE AMAR TELIDJI LAGHOUAT
كـلـيـــــة الـعـلـــــوم
FACULTE DES SCIENCES
DEPARTEMENT DE MATHEMATIQUES ET INFORMATIQUE

Mémoire de MASTER

Domaine : Mathématiques et Informatique
Filière : Informatiques
Option : Systèmes d'Information et de Décision

Par :
HAIMOUD Yamina

THEME

APPLICATION DES ALGORITHMES GENETIQUES POUR L'OPTIMISATION DES REQUÊTES DANS UN SYSTEME DE RECHERCHE D'INFORMATION

*Soutenu publiquement le **04/06/2016**, devant le jury composé de:*

Mr. Tahar ALLAOUI	*M.A.A*	*Président*
Mr. Mustapha BOUAKKAZ	*M.A.A*	*Examinateur*
M^{LLE}. Sara BENKOUIDER	*M.A.A*	*Examinateur*
Mr. Laradj CHELLAMA	*M.A.A*	*Encadreur*

Année Universitaire 2015/2016

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعـلـيـم الـعـالــي و الـبـحـث الـعـلـمـي
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
جـامـعـة عمّــار ثـلـيــجــي بالأغــواط
UNIVERSITE AMAR TELIDJI LAGHOUAT

كلـيــة العـلــوم

قـــســـــم :

مـذكـــــــــرة
للحصول على شهادة المـــاستر في :

مـيـدان
فرع
تخصص

الإسم و اللقب

المـوضـــوع

العنوان هنا

نوقشـــــت علـــنا أمام الجنة المكونة من

السيد	**أستاذ التعليم العالي**	**رئـــــــسا**
	استاذ محاضر	**ممتحنا**
		ممتحنا
		مقررا

السّـنـــــة الجـــــامعية 2015\2014

énétiqueDOptimisationDeRequête4.137 étaheuristiques pour l'optimisation difficile5 éseaux de neurones7 êtebasésur l'algorithmique génétique9 égies d'évolution en environement dynamique11 émantique16 nie/IFT6255/APnie/IFT6255/AP19

UNIVERSITÉ AMMAR TELIDJI LAGHOUATRÉPUBLIQUE
ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA
RECHERCHE SCIENTIFIQUE
UNIVERSITÉ AMMAR TELIDJI LAGHOUAT

FACULTÉ DES SCIENCES
DÉPARTEMENT DE MATHÉMATIQUES ET INFORMATIQUE
Mémoire de MASTER FILIÉRE: INFORMATIQUE
OPTION: SYSTÉME D'INFORMATION ET DÉCISION

par :
HAIMOUD AMINA

THÈME :

Application algorithme génétique pour l'optimisation des requêtes dans un système de recherche d'information

SOUTENU DEVANT LE JURY COMPOSÉ DE :

MME	PRÉSIDENT
MR.	EXAMINATEUR
MR.	EXAMINATRICE
MR. CHELLAMA LAARAJ	ENCADREUR

2015/2016

Résumé

La recherche d'information et la navigation dans les pages web s'avèrent complexes du fait du volume croissant des données indexés.
Les requêtes qui déposent les utilisateurs peut être similaire à des autres requêtes qui sont indexés,elles permettent de retournrer des mauvais documents.
Notre objectif se porte sur l'utilisation des Algorithme Génétique pour l'optimisation des requêtes dans un système de recherche d'information adaptatif aux besoins des utilisateurs.
Parmi les opérations d'un algorithme génétique, on s'intéresse à la fonction de fitness afin d'évaluer l'éxécution des requêtes pour retourner les documents pertinents par les calcules de leurs precision et recall.
Notre expérimentation est faite via l'outil Matlab avec un collection TREC AP88-90

.

Mots-clés : Système de Recherche d'information(SRI), Algorithme Génétique(AG), Optimisation des requêtes, TREC, MATLAB.

Abstract

Information research and browsing in web pages are complex because of the growing volume of indexed data. The queries that users file may be similar to other queries that are indexed, they allow to return bad documents. Our goal is focused on the use of Genetic Algorithm for query optimization in an adaptive information system research to user needs.

Among the operations of a genetic algorithm, we are interested in the function of fitness to assess the execution of queries to return documents relevant by their calculated precision and recall. Our experiment is done via the Matlab tool with TREC collection AP88-90

Keywords : Information Retrieval System , Genetic Algorithm , query optimization,TREC, MATLAB.

Remerciement

Tout d'abord on remercie ***DIEU*** *le tout puissant qui m'a donné la force, la volonté et le courage pour terminer ce modeste travail*

Je tiens à exprimer notre profonde reconnaissance et notre gratitude à mon encadreur ***Mr CHELLAMA LARADJ****, d'abord pour avoir accepter d'encadrer, pour avoir soutenus, dirigés et orientés toute la période de notre projet.*

Nos vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont porté à notre recherche en acceptant d'examiner notre travail Et de l'enrichir par leurs propositions.

Je remercie également tous ceux qui ont contribué de prés ou loin au parachèvement de ce travail, soit par leur savoir scientifique ou par leur amitié.

Dédicace

*louange à **Dieu**, le seul et l'unique*

À mes très chers parents, qui ont fait de moi ce que je su :is

À mes chers frères et sœurs.

À tous mes amis(es)

À tous mes familles HAIMOUD et BEY

À toutes mes amis en témoignage de l'amitié sincère qui nous a liées et des bons moments passés ensemble

À tous ceux qui me sont chers

À tous ceux qui aiment Yamina et ceux qui Yamina aime.

Yamina

Table des matières

Table des figures

Liste des abréviations

AG Algorithme Génétique

RI Recherche d'Information

SRI Système de Recherche d'Information

TREC Text REtrieval Conference

MATLAB MATrix LABoratory

AP88-90 Associated Press

SGML Standard Generalized Markup Language

NIST National Institute of Standards and Technology

DARPA Defense Advanced Reserach Projet Agency.

Introduction générale

Introduction générale

Le volume de données présentés sur le web est en augmentation extrêmement rapide, pour ça l'accès aux données un sujet important dans le monde informatique. On peut dire que le nombre de documents accessibles est de l'ordre de la dizaine de milliards.
Les systèmes de recherche d'information et à partir d'une requête présentée par un utilisateur, ils parcourent leur mémoire pour trouver le plus rapidement possible des documents les plus pertinents parmi les documents qui sont indexés.
Les algorithmes génétiques sont des algorithmes d'optimisation qui sont utilisent dans plusieurs domaines.
L'objectif de notre travail est d'appliquer ces algorithmes pour trouver une solution au problème d'optimisation des requêtes dans un système de recherche d'information.
De manière plus détaillée, nous nous intéréssant au un concept important dans les algorithmes génétiques qui est la fonction d'adaptation.

La structure de notre mémoire s'articule en quatre chapitres :

Le premier chapitre traite des généralités concernant le domaine de recherche d'information, et le système de recherche d'information.
Le second chapitre présente le principe des algorithmes génétiques, nous avons défini le principe et les concepts important des algorithme génétique.
Le troisième chapitre dresse un état de l'art sur les travaux d'application des algorithmes génétiques en recherche d'information.
Le quatrième chapitre est consacré à la présentation et l'implémentation de notre approche .

En fin, nous dressons une conclusion générale et des perspectives d'évolution de ce travail.

Première partie

Recherche d'informatiton et Algorithme Génétique

Chapitre 1

La recherche d'information et le système de recherche d'information

1.1 La recherche d'information

1.1.1 Introduction

La recherche d'information (RI) est une discipline ancienne, elle remonte aux années 1950, est un ensemble des outils et techniques qui permettant de retrouver les documents contenant l'information pertinente à un besoin des utilisateurs dans des corpus, Celui-ci est composé d'un ensemble de documents d'une ou plusieurs bases de données. Aujourd'hui la recherche d'information est une activité quotidienne très exercée par les utilisateurs .

1.1.2 Le processus de recherche d'information

Concérnant la représentation des documents et des requêtes, le système de recherche d'information intègre un ensemble de modèles, qui permet de sélectionner l'information pertinente en réponse au besoin exprimé par l'utilisateur à l'aide d'une requête.Le figure 1.1 présente un processus de recherche d'information.

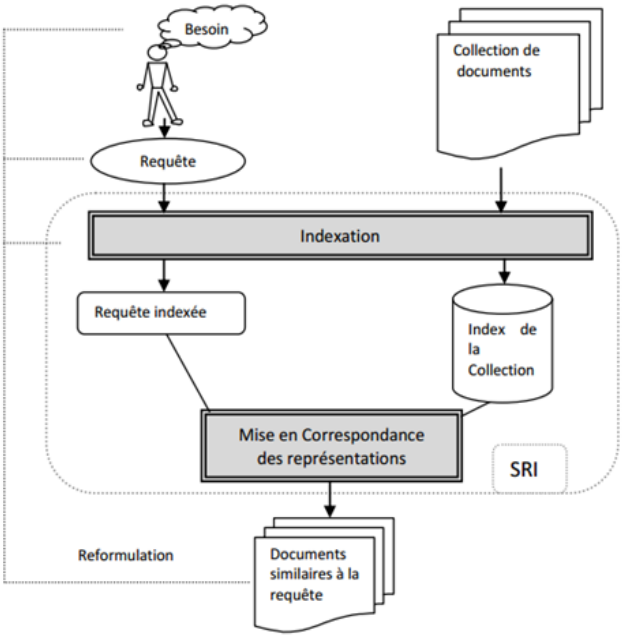

FIGURE 1.1 – Processus de recherche d'information

Cette figure représente le processus de recherche d'information qui permet à l'utilisateur de formuler son besoin d'information sous forme de requête,

celle-ci est alors indexée par le système. Dans le même temps, la collection de documents est également indexée. Grâce à l'index (collection et requête indexées), le système est en mesure de construire les représentations puis de mettre en correspondance la représentation de la requête avec les représentations des documents de la collection. Puis il retourne une liste de documents considérés par le système de recherche d'information comme pertinents par rapport à la requête utilisateur.

L'objectif fondamental d'un processus de recherche d'information est de sélectionner les documents "les plus proches" du besoin en information de l'utilisateur décrit par une requête.

L'indexation

L'indexation est une étape très importante dans le processus de RI. Elle consiste à déterminer et extraire les termes représentatifs du contenu d'un document ou d'une requête, qui couvrent au mieux leur contenu sémantique. La qualité de la recherche dépend en grande partie de la qualité de l'indexation.

Les requêtes

La requête est créée par l'utilisateur, c'est elle qui initie le processus de recherche. Elle traduit un besoin d'information, c'est-à-dire une nécessité ressentie de combler une déficience constatée en information, une lacune ou un défaut. C'est une situation problématique qui amène l'utilisateur à formuler une requête [Schutz et al, 1973].
La requête doit contenir les concepts clés du besoin et les relations entre ces concepts. Elle est issue d'une analyse conceptuelle du besoin d'information qui est effectuée dans l'esprit de l'utilisateur de façon plus ou moins précise. En effet, l'utilisateur fait face à un « problème de vocabulaire » quand il tente de traduire son besoin d'information en une requête.
Une fois la requête exprimée, il est nécessaire de lui donner une forme utilisable par un SRI pour entamer le processus de recherche.

Les documents

Une définition est adoptée pour trois langues (Allemand, Anglais, Français) en 1935 : « Document : Toute base de connaissance, fixée matériellement, susceptible d'être utilisée pour consultation, étude ou preuve. Exemples : manuscrits, imprimés, représentations graphiques ou figurées, objets de collections, etc. »
Le document est une trace d'activité humaine, trace laissée dans l'objectif

d'être interprétée par des personnes souvent différentes du ou des personnes à l'origine de cette même trace.on peut voir le document comme une chose porteuse de sens pour un auditoire donné. Son contenu s'exprime en une forme interprétable pour quelqu'un. Un document peut être un texte, une page WEB, une image, une bande vidéo, etc. Dans notre contexte, nous appelons document toute unité qui peut constituer une réponse à un besoin en information exprimé par un utilisateur.

1.1.3 Les modèles de recherche d'information

Un modèle de RI a pour rôle de fournir une formalisation du processus de RI et un cadre théorique pour la modélisation de la mesure de pertinence. Il existe un grand nombre de modèles de RI textuelle développé dans la littérature. Ces modèles ont en commun le vocabulaire d'indexation basé sur le formalisme mots clés et diffèrent principalement par le modèle d'appariement requête-document. Le vocabulaire d'indexation $V = ti,\ i \in 1, ..., n.$ est constitué de n mots ou racines de mots qui apparaissent dans les documents.
Un modèle de RI est défini par un quadruplet **(D, Q, F, R (q, d)** : où

- **D** est l'ensemble de documents
- **Q** est l'ensemble de requêtes
- **F** est le schéma du modèle théorique de représentation des documents et des requêtes
- **R (q, d)** est la fonction de pertinence du document d à la requête q

Nous présentons dans la suite les principaux modèles de RI : le modèle booléen, le modèle vectoriel et le modèle probabiliste.

Le modèle booléen

Le modèle booléen est le premier modèle de la RI. Il est basé sur la théorie des ensembles. Un document est représenté par l'ensemble des termes qui le composent.
Le modèle booléen peut être expliqué en considérant une requête formée d'un terme comme une définition non ambigüe d'un ensemble de documents. Ainsi la requête retrieval définit simplement l'ensemble de tous les documents indexés avec le terme retrieval. Les requêtes peuvent être composées de plusieurs termes reliés entre eux par des opérateurs de la logique booléenne. Georges Boole a défini trois opérateurs de base : le produit logique AND, la somme logique OR, la différence logique NOT.
Le problème majeur de l'approche booléenne est que les documents qui

répondent à la requête sont retournés dans un ordre quelconque, et sont tous identiquement similaires à la requête.

Le modèle vectoriel

Le modèle vectoriel est un modèle algébrique où l'on représente les documents et les requêtes par des vecteurs dans un espace multidimensionnel dont les dimensions sont les termes issus de l'indexation [Salton, 1983].
D'une manière générale, les résultats de recherche tendent à prouver que les systèmes de recherche vectoriels sont plus performants en terme de précision que les systèmes de recherche booléens [Turtle et al, 1991].

Le modèle probabiliste

Ce modèle est fondé sur le calcul de la probabilité de pertinence d'un document pour une requête. Le principe de base consiste à retrouver des documents qui ont en même temps une forte probabilité d'être pertinents, et une faible probabilité d'être non pertinents. Etant donné une requête utilisateur Q et un document D, il s'agit de calculer la probabilité de pertinence du document pour cette requête.

1.2 Le système de recherche d'information

1.2.1 Définitions d'un SRI

Les systèmes de recherche d'information ont pour but d'offrir des moyens permettant de retourner les informations pertinentes relatives à un besoin en information d'un utilisateur à travers des collections de documents.
Il y a plusieurs définitions d'un SRI, qui sont plus ou moins proches. Tomek Strzalkowski définit un SRI comme suit [Strzalkowski, 1993] :
La tâche typique de la recherche d'information, est de sélectionner des documents dans une base de données, en réponse à une requête de l'usager, et leur rangement par ordre de pertinence .
Salton et Mcgill donnent une définition d'un SRI plus simple mais plus précise et complète [G.Salton and McGill, 1983] : Un SRI traite de la représentation, du stockage, de l'organisation et de l'accés aux éléments de l'information.

1.2.2 La reformulation de requête

La reformulation de requête est un processus permettant de générer une requête plus adéquate à la recherche d'information dans l'environnement du

SRI, que celle initialement formulée par l'utilisateur. Son principe est de modifier la requête de l'utilisateur par ajout de termes significatifs et/ou réestimation de leur poids.
Parmis les technique de reformulation des requêtes :

La reformulation par injection de pertinence

la reformulation par injection de pertinence ou bien (Relevance Feedback), cette méthode permet de faire une modification de la requête initiale, sur la base des jugements de pertinence de l'utilisateur sur les documents restitués par le système.
On utilise la requête initiale pour amorcer la recherche d'information puis exploiter itérativement les jugements de pertinence de l'utilisateur afin d'ajuster la requête par expansion ou repondération.
La nouvelle requête obtenue à chaque itération de feedback, permet de corriger la direction de recherche dans le fond documentaire, et ce, dans le sens des documents pertinents.

1.3 Le problème d'optimisation

L'optimisation est une branche des mathématiques et de l'informatique en tant que disciplines, cherchant à modéliser, à analyser et à résoudre analytiquement ou numériquement les problèmes qui consistent à déterminer quelles sont la ou les solution(s) satisfaisant un objectif quantitatif tout en respectant d'éventuelles contraintes.
Les problèmes d'optimisation occupent actuellement une place très importante dans le domaine de la recherche.
La résolution d'un problème d'optimisation consiste à explorer un espace de recherche afin de maximiser ou minimiser une fonction donnée.Les complexités relatives, en taille ou en structure de l'espace de recherche et de la fonction à maximiser conduisent à utiliser une méthodes de résolution. On peut dire qu'une méthode est adaptée à un espace de recherche complexe et large nécessite plutôt une méthode de recherche stochastique comme les Algorithme Génétique [Berro 2001]

1.3.1 Les différentes méthodes

Il existe plusieurs méthodes comme (le recuit simulé, les algorithmes de colonies de fourmis, les algorithmes génétiques...)pour résourdre des problémes difficile, ces méthodes se basent généralement sur des phénomènes physiques,

biologiques, socio-psychologiques...
Dans notre travaille en utilisent les algorithmes génétiques

1.3.2 Les méthodes d'optimisation combinatoire

Les méthodes d'optimisation peuvent être réparties en deux grandes classes de méthodes pour la résolution des problèmes :

1. Les méthodes exactes,
2. Les méthodes approchées,

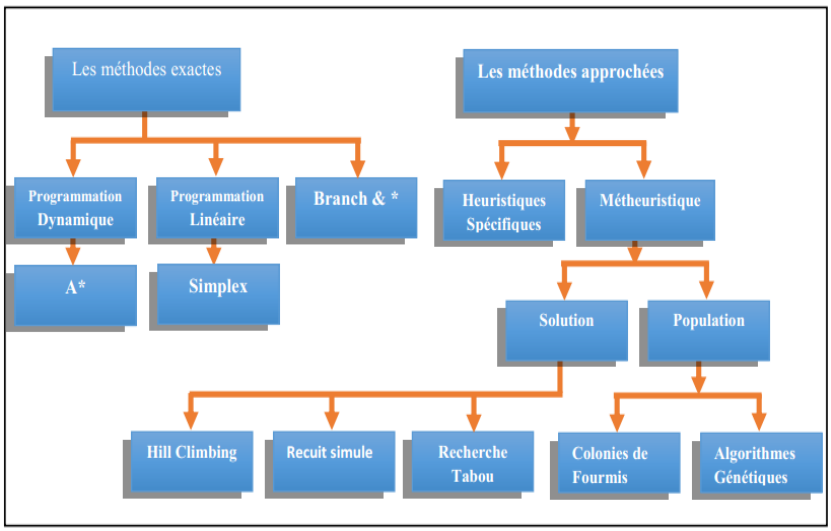

FIGURE 1.2 – Classification des méthodes d'optimisation combinatoire

1.3.3 Les méthodes approchées ou heuristique

Une méthode heuristique ou approchée est une méthode d'optimisation qui a pour but de trouver une solution réalisable de la fonction objectif en un temps raisonnable, mais sans garantie d'optimalité. L'avantage principale de ces méthodes est qu'elles peuvent s'appliquer à n'importe quelle classe de problèmes, faciles ou très difficiles, D'un autre côté les algorithmes d'optimisation tels que les algorithmes de recuit simulé, les algorithmes tabous et les algorithmes génétiques ont démontré leurs robustesses et efficacités face à plusieurs problèmes d'optimisation combinatoires.
Pour notre part, nous nous intéressons aux l'étude des algorithmes génétiques.

Chapitre 2

Les Algorithmes Génétiques

2.1 Introduction

Un problème d'optimisation est défini par un éspace d'état une ou plusieurs fonction(s) objectif(s) et ensemble de contraintes.
L'éspace d'état est défini par l'ensemble des domaines de définition des variables du problème.
Dans la plupart des problème, cet éspace est fini car la méthode de résolution utilisé nécessite un éspace de travail restreint. Par exemple les algorithmes génétiques... .

Parmis tous les types d'algorithmes existants, certains ont la particularité de s'inspirer de l'évolution des espèces dans leur cadre naturel. Ce sont les **algorithmes génétiques**.

Un algorithme génétique va reproduire ce modèle d'évolution dans le but de trouver des solutions pour un problème donné.
Les algorithmes génétiques [Holland, 1975], et plus généralement les algorithmes évolutionnaires, se basent sur les grands principes rencontrés dans la nature et notamment celui de l'évolution des espèces et de la sélection naturelle énoncés par Charles Darwin [Darwin, 1859].
Les algorithmes génétiques sont des algorithmes d'optimisation s'appuyant sur les principes d'évolution naturelle et de sélection des espèces comprenant des croisements entre individus, des mutations apparaissant aléatoirement au sein d'une population et une sélection des individus les mieux adaptés à l'environnement [Holland, 1975]. Face à un problème pour lequel il existe un grand nombre de solutions, l'AG va explorer l'espace des solutions en se laissant guider par les principes décrits précédemment.

2.2 Le but des AG's

Le but des Algorithmes Génétiques est de déterminer les extrêmes d'une fonction $f : X \mapsto R$, où X est un ensemble quelconque appelé espace de recherche,
et f est appelée fonction d'adaptation ou fonction d'évaluation ou encore fonction fitness. La fonction agit comme une «boite noire»pour l'AG. Aussi des problèmes très complexes peuvent être approchés par programmation génétique sans avoir de compréhension particulière du problème.

2.3 Fonctionement d'un Algorithme génétique

Le figure 2.1 présente un organigramme qui illustre le fonctionement l'algorithme génétique
.les étapes important sont, l'étape d'initialisation est générartion de la population initial qui est générée aléatoirement.Le second étape est l'évaluation qui est calculée par la fonction de fitness,elle mesure son adaptation à un environnement donné.L'étape suivante est la reproduction, consiste d'appliquer les opérateurs (Selection,Croisement, mutation) .
Retour à la phase d'évaluation jusqu'à la vérification du critère d'arrêt de l'algorithme.

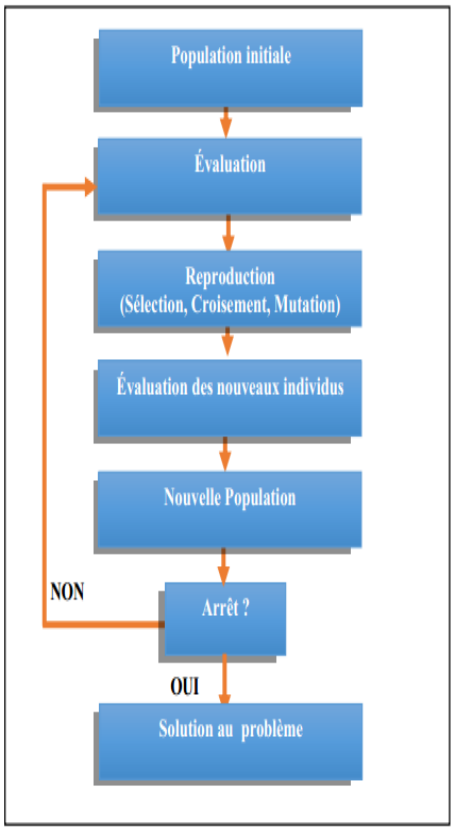

FIGURE 2.1 – Fonctionement d'Algorithme génétique

2.4 Application

Les applications des Algorithmes Génétiques sont multiples :

- Optimisation de fonctions numériques difficiles (discontinues...),
- Traitement d'image (alignement de photos satellites),
- Optimisation d'emplois du temps
- Optimisation de design
- Contrôle de systèmes industriels,
- Optimiser des réseaux (câbles, fibres optiques, mais aussi eau, gaz...),
- Optimiser des antennes...
- Apprentissage des réseaux de neurones [Renders, 1995],

plus généralement les algorithmes génétiques sont désormais largement appliqués en science et ingénierie des algorithmes adaptatifs pour résoudre des problèmes pratiques.

2.5 Les concepts importants des Algorithmes Génétiques

Les Algorithmes Génétiques ont pris les concepts importants de cette dernière. Ces algorithmes ont pris :

Les chromosomes (individus) : sont les éléments à partir des quels sont élaborés les solutions à un problème posé(représente une solution potentielle).

Géne : bit ou ensemble de bits codant une information.

Fitness (coût) : fonction à optimiser, mesure d'efficacité des individus solutions, régissant les transformations génétiques appliquées.et aussi appelé une (fonction d'adaptation).

Population : ensemble d'individus d'une méme génération(espace de recherche).

Le figure 2.2 représente les concepts de base d'un algorithme génétique

2.6 Caractéristiques des Algorithmes Génétiques

2.6.1 Le codage

Le codage est une fonction qui permet de passer de la donnée réelle du problème traité à la donnée utilisée par l'algorithme génétique. Le choix du

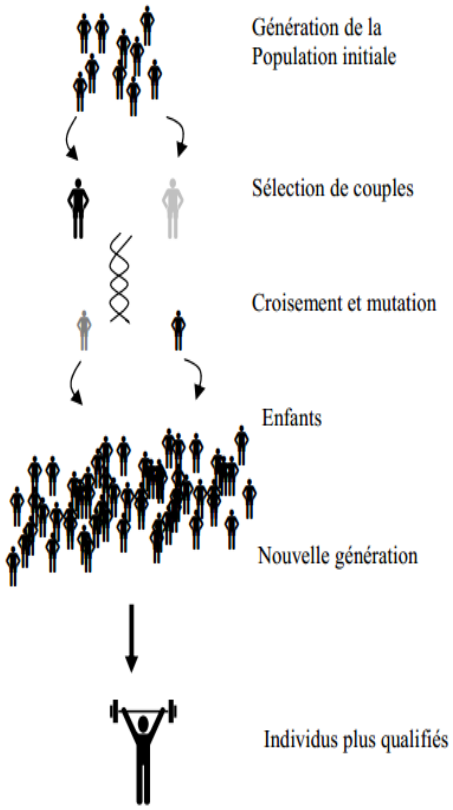

FIGURE 2.2 – Concepts de base d'un Algorithme Génétique

codage est l'élément le plus important dans la conception de l'algorithme puisqu'il permet d'une part de représenter les données, les paramètres et les solutions et d'autre part il influe sur la mise en oeuvre des opérations génétique telles que le croisement et la mutation qui influent directement sur le bon déroulement de l'algorithme génétique et de leur convergence vers la bonne solution.
Généralement on a trois types de codage les plus utilisés :

Représentation binaire

Chaque gène dispose du même alphabet binaire 0, 1. Un gène est alors représenté par un entier, les chromosomes, qui sont des suites de gènes sont représentés par des tableaux de gènes et les individus de l'espace de recherche sont représentés par des tableaux de chromosomes.

Représentation avec réels

Contrairement au codage binaire, un gène est représenté par une suite de bits (un bits dans le code binaire) qui est associé à un réel. Ce type de codage peut être utile notamment dans le cas où l'on recherche le maximum d'une fonction réelle.

Représentation à l'aide d'arbres syntaxique

Ce type de codage utilise une structure arborescente (une racine de laquelle peuvent être issus un ou plusieurs fils eux mêmes des arbres). Un arbre syntaxique est un arbre contenant deux types de noeuds [Philippe Laublet et Chantal Reynaud Jean Charlet, 2005] :

1. les noeuds internes ou « symboles non terminaux »

2. les feuilles ou «symboles terminaux »

Ce type de codage peut être utilisé lorsque la taille du problème ou de la solution n'est pas finie. Son inconvénient est qu'on peut trouver des arbres de solutions de taille importante difficile à analyser.

2.6.2 La fonction d'évaluation

On appelle aussi fonction objectif ou fonction d'adaptation ou fonction fitness, associe une valeur de performance à chaque individu ce qui offre la possibilité de le comparer à d'autres individus et permet à l'algorithme génétique de déterminer qu'un individu sera sélectionné pour être reproduit ou pour déterminer s'il sera remplacé [Eric Taillard Patrick Siarry Johann

Dréo, Alain Petrowski].
La fonction d'adaptation est un élément fondamental lors de la modélisation d'un AG .

2.7 Les opérateurs génétiques

Ces opérateurs sont la base des Algorithmes Génétiques, Les opérateurs qu'on retrouve le plus souvent sont : sélection, croisement et mutation.

2.7.1 Sélection

Qui est le choix des chromosomes de la population à reproduire.Plusieurs méthodes existent pour sélectionner des individus destinés à la reproduction. Il y' a plusieurs méthodes de sélection, les plus utilisées sont :

Sélection par rang Choisir toujours les individus possédant les meilleurs scores.

Sélection par tournoi Choisir aléatoirement deux individus et on compare leur fonction d'adaptation (combattre) et on accepte le plus adapté pour accéder à la génération intermédiaire, et on répète cette opération jusqu'à remplir la génération intermédiaire (N/2 composants). Les individus qui gagnent à chaque fois on peut les copier plusieurs fois ce qui favorisera la pérennité de leurs gènes.[choisir parmi ces paires(deux individus) celui qui a le meilleur score d'adaptation]

Sélection uniforme Choisir aléatoirement sans faire intervenir la valeur d'adaptation.

Roulette de casino C'est la sélection naturelle la plus employée pour l'AG binaire. Chaque chromosome occupe un secteur de roulette dont l'angle est proportionnel à son indice de qualité. Un chromosome est considéré comme bon aura un indice de qualité élevé, un large secteur de roulette et alors il aura plus de chance d'être sélectionné.

Elitisme La stratégie élitiste consiste à conserver le meilleur individu à chaque génération. Ainsi l'élitisme empêche l'individu le plus performant de disparaître au cours de la sélection ou que ses bonnes combinaisons soient affectées par les opérateurs de croisement et de mutation. Après chaque évaluation de la performance des individus à une génération t donnée, le meilleur individu de la génération précédente (t1) est réintroduit dans la population si aucun des individus de la génération t n'est meilleur que lui. Par cette approche, la performance du meilleur individu de la population courante est monotone de génération

en génération. Il apparaît que l'élitisme améliore considérablement les performances de l'algorithme génétique pour certaines classes de problèmes, mais peut les dégrader pour d'autres classes, en augmentant le taux de convergences prématurées.

2.7.2 Croisement

Qui prend une séquence de gènes de chacun de deux chromosomes (dits parents) choisis et les combine pour créer un nouveau chromosome en résultat(dits enfants).

Le croisement à un point Il a été initialement défini pour le codage binaire. Le principe consiste à tirer aléatoire une position pour chaque parent et à échanger les souschaines des parents à partir des positions tirées.

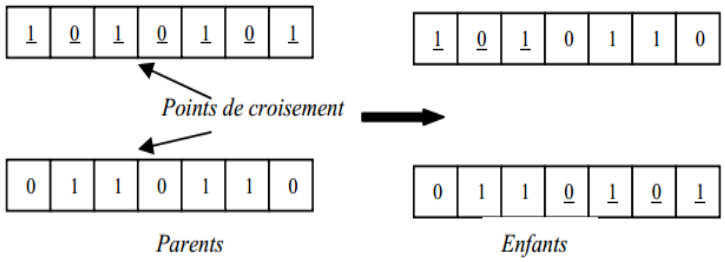

FIGURE 2.3 – Exemple du croisement à un point

Le croisement à deux points Elle reprend le mécanisme de la méthode de croisement à un point en généralisant l'échange à 3 ou 4 sous chaines.

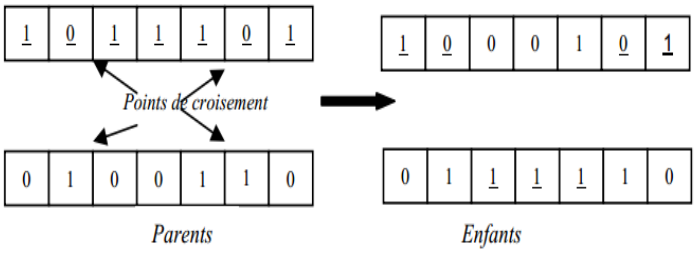

FIGURE 2.4 – Exemple du croisement à deux points

2.7.3 Mutation

Est une opérateur de modification de structure d'individus.Il permet de protège les algorithmes génétiques des pertes prématurées d'informations pertinentes. et elle permet d'introduire une certaine information dans la population, qui a pu être perdue lors de l'opération de croisement. Ainsi elle participe au maintien de la diversité, utile à une bonne exploration du domaine de recherche.

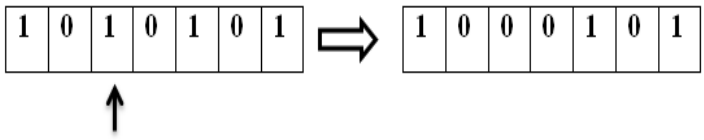

FIGURE 2.5 – Exemple de la mutation

2.8 Les avantages des AG's

Un des grands avantages des Algorithmes Génétiques est qu'ils autorisent la prise en compte de plusieurs critères simultanément, et qu'ils parviennent à trouver de bonnes solutions sur des problèmes très complexes. L'avantage principal des algorithmes génétiques par rapport aux autres techniques d'optimisation combinatoire consiste en une combinaison de l'exploration de l'espace de recherche, et de l'exploitation des meilleures solutions disponibles à un moment donné. Ils doivent simplement déterminer entre deux solutions quelle est la meilleure, afin d'opérer leurs sélections. Leur utilisation se développe dans des domaines aussi divers que l'économie, la bioinformatique ou la vérification formelle.

2.9 Les inconvénient des AG's

L'inconvénient majeur des Algorithmes Génétiques est le coût d'exécution important par rapport à d'autres méta-heuristiques. Les algorithmes génétiques nécessitent de nombreux calculs, en particulier au niveau de la fonction d'évaluation. Mais avec les capacités calculatoires des ordinateurs récents,

ce problème a perdu de son importance. D'un autre côté, l'ajustement d'un algorithme génétique est délicat : des paramètres comme la taille de la population ou le taux de mutation sont parfois difficiles à déterminer

2.10 Conclusion

Ce chapitre a porté essentiellement sur les Algorithmes Génétiques et le opérateurs , ainsi les concepte important, le fonctionement de ces algorithmes .

Les Algorithmes Génétiques sont les approches métaheuristique les plus répandues pour résoudre des problèmes difficiles d'optimisations et de recherche. Leur efficacité est déterminée par les opérateurs génétiques qui sont utilisés et par la fonction d'évaluation.

Deuxième partie

Les Algorithmes Génétiques à la Recherche d'Information

Chapitre 3

Algorithmes Génétiques en recherche d'information

3.1 Introduction

Plusieurs travaux ont été publiés récemment dans le domaine de la recherche d'information. Jhon Holland qui mit les bases des premières applications des algorithmes génétiques en 1962, pour notre mémoire nous nous intéressons au domaine de systéme de recherche d'information. Vu sous l'angle de l'optimisation, les techniques de recherche d'information ciblent trois principaux objectifs [Lynda Tamine] :

1. **Représentation optimale des documents :** Consiste à couvrir de manière fidèle la sémantique véhiculée par un document en considérant le contenu de la collection.
2. **Représentation optimale des requêtes :** Consiste à traduire l'intégralité de la sémantique véhiculée par la requête en considérant le véritable besoin en informations de l'utilisateur ainsi que le contenu de la collection interrogée.
3. **Formalisation optimale de la fonction pertinence :** Cette dernière traduit une combinaison formelle de critères permettant d'estimer la pertinence d'un document relativement à une requête.

La recherche d'information tente de trouver les documents qui répondent à un besoin exprimé par un usager, et ce parmi une grande collection de documents.Soltan présente des techniques qui permettent d'évaluer le poids d'un terme dans un document appartenant à une collection [Gerard Salton]. La recherche d'information dans une collection de données dépend de l'indexation des documents, l'analyse sémantique de la requête et le calcul de la similarité entre la requête exprimée par l'utilisateur et les documents de la collection pour permettre par la suite de classer les documents par leurs valeurs de similarité à la requête. La difficulté réside dans la recherche de la sémantique associée aux documents et à la requête.

3.1.1 Reformulation de requête

La reformulation de requête est un mécanisme de modification de requête par expansion et/ou repondération de la requête initiale en utilisant des critères de choix définis sans intervention de l'utilisateur. Ce type de reformulation peut être défini dans un contexte global, basé sur le thésaurus, ou alors local, basé sur les résultats de recherche en cours.
La reformulation basée sur le contexte global est essentiellement basée sur l'ajout et repondération de termes issus d'un thésaurus manuel en utilisant des calculs de poids de similarité, cooccurrence et relations contextuelles entre termes ou une combinaison de divers types de relations.

La reformulation basée sur le contexte local, utilise des informations issues de la recherche en cours : documents retrouvés, termes et poids associés.
Les deux types de reformulation diffèrent essentiellement par la source d'information utilisée quant à la dérivation de l'association sémantique entre termes ou entre termes et documents, mais utilisent toutes les deux des fonctions caractéristiques pour la sélection et repondération des nouveaux termes de la requête.
Le présent section a pour objectif de présenter les principaux travaux d'application des algorithme génétique à la Recherche d'Information

3.2 Les travaux

3.2.1 Application des Algorithmes Génétiques pour l'ptimisation des requêtes

Il existe plusieurs travaux dans le cadre de l'optimisation des requêts, parmi ces travaux :

Yang et Korfhage 1993 sont développèrent un AG pour une optimisation de requête par réestimation des poids d'indexation sans induire une expansion. Un individu requête est représenté comme une liste pondérée de termes d'indexation. Les générations de requêtes sont renouvellées par application :
- d'une fonction d'adaptation basée sur la formule :

$$Fitness(q)=\alpha R_{p}^{(q)}-\beta R_{Np}^{(q)}$$

Où :

q : Individu requête
Rp(q) : Nombre de documents pertinents retrouvés
RNp(q) : Nombre de documents non pertinents retrouvés
- d'une sélection basée sur un échantillonnage stochastique ,
- d'un croisement à deux points et d'une mutation classique.

Kraft et al 1995 sont appliquent les techniques de programmation génétique dans le but d'optimiser la représentation des requêtes dans le modèle booléen. Un individu requête est représenté sur la base du modèle génétique de Koza [Koza, 1991]. Leurs premières expérimentations ont montré la faisabilité de l'approche pour dériver des requêtes qui accroissent les performances du système en termes de rappel et précision.

3.2.2 Application des algorithmes Génétiques à La recherche d'information dans le WEB

Conçernant la recherche d'information dans le web, plusieurs travaux sont présenté, parmi ces travaux :

Nick et Themis, 2001 proposent un système qui s'appelle Webnaut(qui est un bon représentant de ce type d'agents personnels.Il combine un méta-moteur interrogeant les moteurs de recherche classiques (Google, ...) avec un algorithme génétique permettant de générer de nouvelles requêtes afin de trouver de meilleurs résultats).
Dans cette algorithme génétique deux opérateurs génétique sont utilisés :Un opérateur de croisement à un point de coupure agissant sur deux individus, et un opérateur de mutation.Cet opérateur consiste à choisir aléatoirement un élément d'un vecteur représentant un individu et à le remplacer par un autre, tiré au hasard dans le cas de la population de vecteurs de booléens, et issu des documents initiaux dans l'autre population..

Vallin et Coello, 2003 leur but est également de créer un agent chargé d'assister les utilisateurs dans leurs problèmes de recherche d'information. La population prise en compte dans l'algorithme génétique est composée d'individus représentés par un vecteur de termes représentant un document correspondant à la requête. La fonction de fitness utilisée calcule une valeur représentant l'adaptation de cette requête à l'intérêt de l'utilisateur.
Dans cette AG ils utilisent Un opérateur de croisement à deux points est

utilisé afin de mélanger les gènes correspondant aux termes représentés dans les vecteurs de documents, dimensionnés de manière similaire.et utilisent l'opérateur de mutation qui consiste à sélectionner aléatoirement des poids de termes dans les vecteurs de documents et à les modifier en fixant une valeur choisie au hasard entre 0 et 1. L'utilisateur peut renforcer ou atténuer l'intérêt d'un document en lui donnant un score respectivement positif ou négatif. Les poids des termes dans les vecteurs représentant les individus sont alors mis à jour en conséquence et la valeur de fitness calculée est augmentée ou diminuée de la même manière.

3.3 Conclusion

Nous avons présenté dans ce chapitre les principales travaux conçernat l'application des algorithme génétique à la recherche d'infomation et l'optimisation des requêtes.
Le chapitre suivant est consacré à la présentation et implimentation de notre approche qui basé sur l'application des Algorithme Génétique au ce domaine.

Chapitre 4

Présentation et implémentation de notre Approche

4.1 Introduction

Les algorithmes génétiques ont montré leur efficacité dans la résolution de nombreux problèmes et notamment dans les problèmes d'optimisation.Sont des techniques importantes dans les recherches aléatoires pour la meilleure solution parmi un groupe de solutions dans les données disponibles . La recherche d'information dans des collection des documents dépend de l'idexation de cette documents dans ces collection. On calcule la similaité entre la requêtes exprimée par l'utilisateurs et les documents de collection .

Nous décrivons dans cette chapitre, notre approche pour l'optimisation de requête dans un système de recherche d'information.

4.2 Description de l'AG d'optimisation de requête

4.2.1 L'Algorithme Génétique d'optimisation de requête

Début
t :=0
Répéter
générer la population initial
Effectuer le codage
évaluation de la fonction de fitness
Fait
Appliquer les opérateurs génétiques
t :=t+1
Jusqu'à *arrêt*
Fin

FIGURE 4.1 – L'Algorithme Génétique d'optimisation de requête

4.2.2 Codage des individus

Le codage des individus est une étape de modélisation fondamentale dans les algorithmes génétiques,
Un individu requête est représenté comme suit :
Où :
Qu(s) : Individu requête numéro 'u' de la population à la génération 's'
t1, t2, t3 ..., tT : Liste de termes d'indexation
qui : Poids du terme ti dans la requête individu Qu(s)

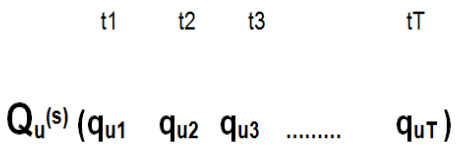

FIGURE 4.2 – Code d'un individu requête

Le poids de terme ti dans la requête individu Qu(s), qui est calculé selon la formule suivante :

$$q_{ij} = \begin{cases} \dfrac{nq*qtf}{nq-qtf} & \text{si } (nq>qtf) \\ & \text{sinon } qtf \end{cases}$$

Où :
qtf : Fréquence d'un terme dans une requête,
nq : Nombre de termes dans la requête

4.2.3 La fonction d'adaptation(Fitness)

L'utilisation d'algorithme génétique nécessite une fonction d'évaluation, cette fonction permet de mesurer la performance d'un individu dans la résolution d'un problème posé.
Pour notre travail La fonction d'adaptation ou bien la fonction de Fitness sera considéré comme des fonctions de precision et de recall.
lorsque l'utilisateur lance leur requête, il attend un nombre de réponse. Pour evaluer ou bien mesurer l'efficacité d'une technique de recherche d'informations par l'utilisateurs, en utilisant deux mesures distinctes qui sont le precision et recall.

precision : est le nombre de documents pertinents retrouvés rapporté au nombre de documents total proposé par le système pour une requête donnée.

le precision est calculé selon la formule suivante :

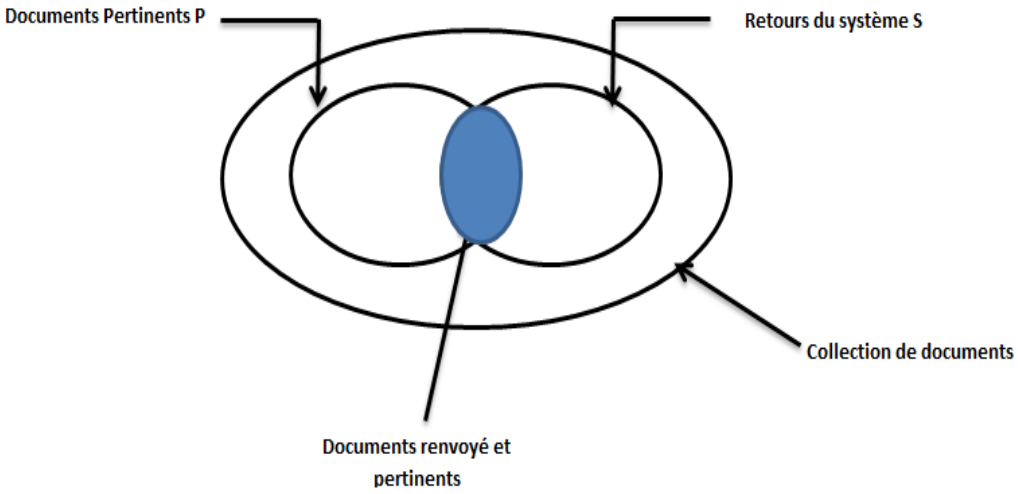

FIGURE 4.3 – Recall et precision

On prend que :

P = les documents pertinents
S = la réponse du système

$$\boldsymbol{PrecisionFitness} = \alpha * \sum[\mathbf{P} \cap \mathbf{S}] / \sum[\mathbf{S}]$$

Recall : est défini par le nombre de documents pertinents retrouvés au regard du nombre de documents pertinents que possède la base de données.
Le recall est calculé selon la formule suivante :

$$\boldsymbol{RecallFitness} = \sum[\mathbf{P} \cap \mathbf{S}] / \sum[\mathbf{P}]$$

Où α est le poids arbitraire. Il est ajoutée spécialement à la fonction de precisionFitness.

4.2.4 Les opérateurs génétique

Selection

L'opérateur de sélection permet de sélectionner les meilleurs individus dans la population pour participer dans la génération de la prochaine population. Nous avons utilisé la sélection par élitisme, où on sélectionne les parents qui ont le minimum de Fitness. ce type de selection retourne la une meilleurs individus de la population.

Croisement

L'opérateur de croisement permet à deux individus d'échanger leurs gènes en vue de créer de nouveaux individus plus intéressants. Le croisement est appliqué sur deux individus pères choisis par l'opérateur de sélection.
Dans notre approche on utilise un croisement à deux points, dans le but d'exploiter au mieux la distribution des termes dans les documents pertinents.

Mutation

L'opérateur de mutation permet de modifier occasionnellement des gènes d'un individu pour permettre d'explorer certaines zones dans la codification des individus où le croisement ne peut pas explorer.

4.3 Implémentation

Nous avons implémenté notre approche sous MATLAB (MATrix LABoratory)

4.3.1 Les outils utilisés

MATLAB

Est un logiciel pour le calcul scientifique, orienté vers les vecteurs et les listes de données. Matleb est un langage interprété, chaque ligne d'un programme matlab est lue, interprétée et exécutée. L'avantage est qu'il est très simple et très rapide à programmer, offrant une grand tolérance (syntaxe simple, pas de définition de types, ...etc), ce qui permet un gain appréciable en temps de mise au point.

Optimtool

Est un application d'optimisation contient plusieurs algorithme ou bien des solver pour faire l'optimisation d'un tel fonction, et des options d'optimisation, et exécuter des problèmes données.

Le solver optimtool commence l'application d'optimisation avec le solveur spécifié ;

conçernant nos travaille on utilise le solver (ga Genetic Algorithm)

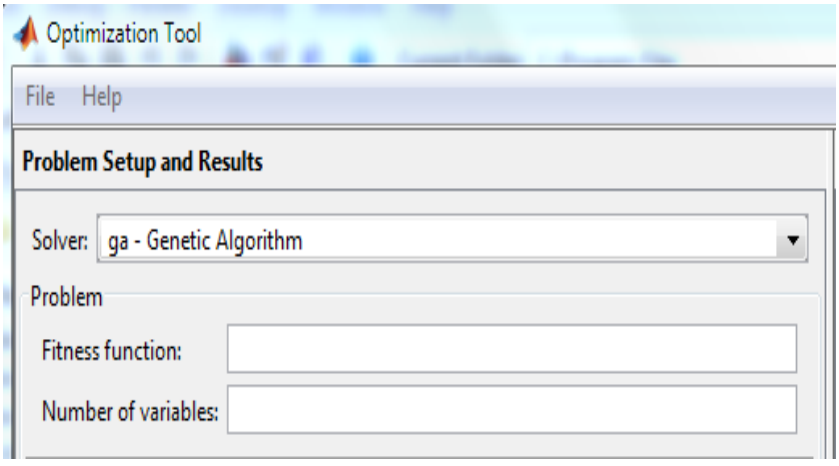

FIGURE 4.4 – la fenêtre du solver ga

Les document sont stocké dans une base de donnée à partir d'un fichier SGML dans un dossier spécifié.
La base de données permet de MATLAB pour rechercher du contenu dans les fichiers SGML.

Syntaxe : builddocsearchdb (dossier)
builddocsearchdb permet de construire la base de données de documentation consultable.

4.3.2 Exprémentation

Nous avons appliquent l'algorithme génétique sur un envirenement de test pour voire l'influence de l'Algorithme Génétique sur un requête.Nos requêtes sont des requêtes booléenes avec l'utilisation d'un modéle booléen d'une collection de documents TREC(**T**ext **RE**trieval **C**onference).
Ces exprémentation visent à appliqué les Algorithme Génétique pour mesurer la perfermance de réponse du système, selon les deux fonction *PrecisionFitness et RecallFitness*

Dans nos exprémentation on évaluer l'impact des valeurs de probabilités de croisement et probabilité de mutation sur les résultats de recherche.
A cet effet, nous avons fait nos test avec ces valeurs de probabilités de croisement (0.6 , 0.7 , 0.8) et les valeurs de probabilité de mutation(0.25 , 0.30 , 0.30).

Collection de test

Pour arriver à une telle évaluation, on doit connaître d'abord les réponses idéales de l'utilisateur. Ainsi, l'évaluation d'un système se fait à l'aide d'un collection de test.
Pour qu'un collection de test soit significatif, il faut qu'il possède un nombre de documents assez élevé.Différentes collections de test sont utilisées en recherche d'information.Parmi elles nous utilisons la collection TREC.

Le projet **TREC** est un programme international initié au début des années 90 par le NIST (National Institute of Standards and Technology) et du DARPA (Defense Advanced Reserach Projet Agency). Ce programme offre des moyens homogènes d'évaluation des systèmes de recherche d'information. Il est devenu la référence en recherche d'information pour diverses raisons. En effet, il a permis de définir les tâches en recherche d'information et de

construire de larges collections de test.
Un document TREC est généralement présenté sous le format SGML.

Nous avons utilisé une petite collection de test TREC AP88-90 [19] , un ensemble d'article de nouvelles publiées par Associated Press en 1988-1990.

Les requêtes sont issues des topics numérotées de 101 jusqu'à 150 de la collection TREC ; nous avons utilisé juste les champs titre qui contient des termes.
Le figure 4.5 présente la Structure de la collection de test TREC AP88, le figure 4.6 présente la structure d'un document TREC identifié AP880100, et le figure 4.7 présente la structure du requête numéro 101,102, et 150 dans la collection TREC.

Le contenu contextuel	*Taille (Mb)*	*nombre de document*	*Nombre moyen de* termes *par document*
AP (Associated Press), 1988	**237**	**79919**	**438**

FIGURE 4.5 – Structure de la collection de test TREC AP88

```
<doc>
<docno> AP880100 </docno>
<author>  Associated press </author>
<dateline> NewYork </dateline>
<text>
African National Congress, ANC, Nelson Mandela,
Oliver Tambo
</text>
<>/doc
```

FIGURE 4.6 – Structure d'un document TREC identifié AP880100

```
<top>
<num> Number: 101
<dom> Domain: Science and Technology
<title> Topic: Design of the "Star Wars" Anti-missile Defense
System
.
.
.
</top>

<top>
<num> Number: 102
<dom> Domain: Science and Technology
<title> Topic: Laser Research Applicable to the U.S.'s
.
.
.
</top>









<top>
<num> Number: 150
<dom> Domain: U.S. Politics
<title> Topic: U.S. Political Campaign Financing
.
.
.
</top>
```

FIGURE 4.7 – structure du requête numéro 101,102 et 150 dans la collection TREC

Discussion

La comparaison des réponses d'un système pour une requête avec les réponses idéales nous permet d'évaluer les deux métriques suivantes : precision et recall.

Les deux métriques ne sont pas indépendantes. Il y a une forte relation entre elles : quand l'une augmente, l'autre diminue.
Le figure 4.8 présente le courbe de precision et recall

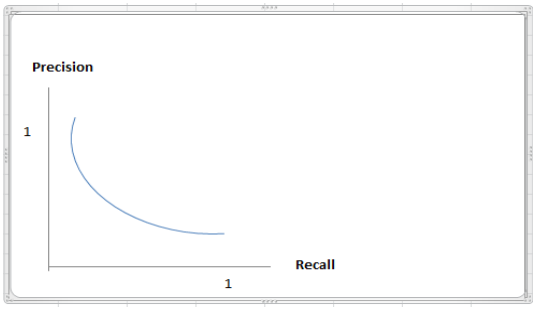

FIGURE 4.8 – Courbe de precision et recall

Nous avons fait plus d'une expérience en combinant les différentes valeurs de taux de mutation et taux de croisement, les résultats obtenus sont montrés dans le tableau 4.1 ,
Dans nos expérimentations le valeur de α est fixé $\alpha = 0,2$

Dans notre implémentation, les valeurs de la fonction de precisionFitness peut être supérieure à 1, où la valeur maximale de précision est de 1,12 qui vient de α ainsi nous ne pouvons pas l'interpréter comme la probabilité.

Tc/Tm	Precision	Recall
60/25	*1.04*	*1.25*
70/30	*1.12*	*0.91*
80/30	*0.93*	*1.02*

TABLE 4.1 – le résultat du taux de croisement et taux de mutation sur la recherche dans la collection

Les résultats obtenus montrant que le derniér test de (Tc/Tm= 80/30) donnera un bon résultat, que les valeur de precision et mieux que les deux premier test.

Le but de nos expérimentations est d'appliqué un Algorithme Génétique pour optimiser et montrer que le système de recherche d'information retourne un bon résultat aux requêtes utilisateurs .
Nous conclurons que les valeurs de precision et recall jouent un rôle important pour l'évaluation de la recherche d'information dans cette collection.
Un système dont la courbe dépasse (c'est-à-dire qu'elle se situe en haut à droite) celle d'un autre est considéré comme un meilleur système, et donnera des résultat plus pertinent aux requêtes utilisateurs.

4.4 Conclusion

Dans ce chapitre, nous avons proposé une approche pour l'optimisation des requêtes. On prend que les deux fonctions(precision et recall) comme des fonctions d'adaptation(fitness) qui permettent de mesuré la perfermance d'un individus dans la résolution de notre probléme d'optimisation des requêtes.Avec l'appplication des opérateurs génétiques.

Conclusion générale

Conclusion

Dans ce mémoire, nous nous sommes intéressés à l'optimisation des requêtes. Pour cela, nous avons proposé une approche d'optimisation de requêtes basée sur les concepts de la génétique.
Nous avons implémenté notre approche sous matlab,et pour nos exprimentations nous nous utilisent la collection de test TREC, pour evaluent notre approche .
De nombreux travaux ont été utilisés dans la littérature pour l'optimisation des reqûtes dans un système de recherche d'information.
Dans ce contexte de travail, nous nous sommes intéressés à la mise en œuvre d'un AG d'optimisation de requête. Les AG's sont des procédés d'optimisation qui présentent des propriétés fort intéressantes : exploration parallèle et efficace d'espaces complexes, construction graduelle de solutions partielles, recherche coopérative. Ces algorithmes sont exploités en vue de construire, dans l'espace défini dans la collection de documents, la structure de la requête optimale, permettant de rappeler le maximum de documents pertinents associés au besoin en information de l'utilisateur.

conçernant notre approche en utilise la fonction de recall et precision comme des fonction d'évaluation(fitness).

Perspectives

Ce travail peut être amélioré, en premier temps par l'évaluation sur d'autre collection qui est caractérisé par autres documents, et nombre diffèrent des termes, et longueur des requêtes...
En second par le changement des opérateurs de l'algorithme génétique.

Bibliographie

[1] [Strzalkowski, 1993] *:Natural language processing in large-scale text retrieval tasks.* In Text REtrieval Conference (TREC-1),1993

[2] [G.Salton and McGill, 1983] *:Introduction to Modern Information Retrieval.* New York, 1983

[3] [G.Salton, 1989]*Automatic Text Processing, The Transformation, Analysis, and Retrieval of Information by Computer* Addison-Wesley, Reading (MA).

[4] [john Holland, 1975] *:Adaptation in natural artificial systems.* University of Michigan Press, Ann Arbor, 1975.

[5] [Eric Taillard Patrick Siarry Johann Dréo, Alain Petrowski] *Métaheuristiques pour l'optimisation difficile.* Eyrolles, Juillet 2003.

[6] [C.Darwin, 1859] *:The Origin of Species by Means of Natural Selection.* Mentor Reprint, 1958, NY, 1859.

[7] [Jean-Michel RENDERS] *:AG et réseaux de neurones.* 1995.

[8] [H. Chen, 1995] *:Machine learning for information retrieval.* neural networks, symbolic learning, and genetic algorithms. Journal of the American Society for Information Science 46 3 (1995)

[9] [Lynda Tamine] *:Reformulation automatique de requête basé sur l'algorithmique génétique.* Actes du congrès Inforsid, Toulouse juin 1997.

[10] [Gerard Salton] *:Automatic text processing : the transformation, analysis, and retrieval of information by computer.* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989.

[11] [Berro 2001] *Optimisation multiobjet et stratégies d'évolution en environement dynamique*Thése du doctorat, spécialité informatique, université des science sociale, Toulouse 1, 2001

[12] [Schutz et al, 1973] *Structures of the Life World*Thése du doctorat, spécialité informatique,Northwestern University Press, Evanston,New York, Sept. 1973

[13] [Turtle et al, 1991] *Evaluation of an inference network-based retrieval model.*ACM Transactions on Information Systems, 1991.

[14] [Nick et Themis, 2001] *Web search using a genetic algorithm*IEEE Internet Computing, 5(2) :18–26, 2001.

[15] [Yang Korfhage 1993] *Query Optimisation in Information Retrieval Using Genetic Algorithms* ICGA, 1993

[16] [Philippe Laublet et Chantal Reynaud Jean Charlet, 2005] *Le Web sémantique* Cépaduès-Èditions, Avril 2005.

[17] [Kraft al, 1995] *Applying Genetic Algorithms to Information Rtrieval System Via Relevance Feedback* In Bosc and Kacprzyk J eds, Fuzziness in Databse Management Systems Studies in Fuzziness Series, Physica Verlag, Heidelberg, Germany

[18] [Vallin et Coello, 2003] *An agent for web information dissemination based on a genetic algorithm*In IEEE International Conference on Systems, Man and Cybernetics - IEEE SMC'03, pages 3834–3839, Hyatt Regency, Washington, D.C., USA, 5 - 8 October 2003. IEEE Press.

[19] *http ://www.iro.umontreal.ca/ nie/IFT6255/AP*