

**Projet De fin d’étude**

**(Option SIT)**

|  |
| --- |
| **Conception et réalisation d’un éditeur de textes Coraniques authentiques** |

**Réalisé par : Encadré par :**

* HACENE Abdessamed Dr CHALLAL Yacine
* TALEB Mustapha Younes Dr MENACER Mohamed

**Année : 2014/2015**

1. Sommaire

Résumé

Abréviations

Liste des tableaux

Liste des figures

Introduction générale (Noor research center et notre projet)

Problématique (le Coran et l’édition de texte)

1. problématique
2. La solution proposée

Objectifs du projet (vérification, validation, suggestion, correction, web services, plugin, sécurité)

# Partie 1 : état de l’art

## Chapitre I : Le Coran

1. Introduction

Le Coran est la parole d’Allah et le livre sacré des musulmans depuis le 7 ème siècle, il est la source de plusieurs sciences notamment L’exégèse (tafssir en arabe), la traduction du Coran et le domaine des miracles scientifiques du Coran

Avec le développement des technologies de l’information il est devenu primordiale de développer des outils permettant de faciliter les sciences Coranique.

Dans ce chapitre nous allons parler de la définition du saint Coran, son historique, ses caractéristiques, puis on termine par la relation entre l’informatique, le Coran et les sciences Coranique.

1. Définitions
   1. Le Coran

Dans la langue arabe Coran signifie la récitation ou la lecture, c’est le texte religieux central de l’islam, et le nom de son livre sacré, parole d'Allah. les révélations faites par Dieu (à Lui la puissance et la gloire) au prophète Mohammad (la prière et le salut soient sur lui) par l'intermédiaire de l'ange Gabriel (Gibrîl) écrits dans les mushafs et transmis par générations successives.

Vue son importance, le saint coran possède plusieurs noms notamment Al-Qurān Al-Furkān, Al-kitāb, Al-dhikr, Alwahy et Al-rōuh **[**[**Mahssin 1973**](#_ENREF_30)**]**

* 1. Le Mushaf

C’est un livre qui regroupe la totalité du saint coran, le premier mushaf a été rassemblé dans la période du calife Abu-bakr.

Les mushafs sont écrits avec le scripte Othmani, le premier qui a utilisé ce scripte est le calife Othman, ce scripte est un peu différent du scripte arabe standard **[**[**CHELLI 2010**](#_ENREF_16)**]**

La traduction du Coran ne constitue en aucun cas un mushaf, car le Coran fut révélé par

Allah en arabe dans sa forme comme dans son sens. **[**[**Hadj henni 2008**](#_ENREF_21)**]**

* 1. Le document Coranique

C’est un document qui peut correspondre à la totalité du contenu du Coran (Mushaf) ou bien à un extrait cohérent (ayates successives) du Coran en langue arabe. Ces extraits peuvent être utilisés comme des citations et exemples. **[**[**Hadj henni 2008**](#_ENREF_21)**]**

Les documents Coranique peuvent être de plusieurs types : texte, image, vidéo et son, dans notre projet nous allons nous focaliser sur les documents coraniques de type texte

* 1. Les sciences Coranique

C’est l’ensemble des sciences qui prennent le Coran comme sujet d’étude, ils ont pour objectifs de servir le coran et illustrer ses secrets et miracles, on distingue deux types de sciences Coranique **[**[**الطيار 2008**](#_ENREF_51)**]**

* **Des sciences provenant originalement du Coran :** Qui ne peuvent pas être utilisé dans d’autres sciences comme l’exemple de tafssīr et les causes de révélation.
* **Sciences classiques en commun avec le coran et d’autres sciences** :

Ce sont des sciences dont la majorité s’intéressent aux aspects linguistiques c’est le cas de la conjugaison (الإعراب) et la Rhétorique (البلاغة) et les aspects législatifs, on prend à titre d’exemple El fiqh (الفقه) **[**[**CHELLI 2010**](#_ENREF_16)**].**

D’après les savants, ces sciences sont nombreuses (jusqu’à 47) sont citées par l’Imam Zarkachi, parmi ces sciences nous trouvons :

* Les causes de la révélation
* La connaissance des exemples existants dans le Coran
* La connaissance des liens entre les ayates
* La science des semblants
* La science des ambiguës
* La connaissance de la conjugaison
* Les arcanes des initiaux des sourates
* Les Ayates Mecquoise et Médinoise
* l’abrogeant et de l’abrogé
* Les dialectes de révélation
* Analyse grammaticale du Coran
* Tafssīr du Coran
* La connaissance de l’orthographie

1. Historique du Coran

Depuis sa révélation jusqu’à nos jours le Coran est passé par plusieurs périodes on les résume dans les point suivant :

* 1. Le Coran à l’époque du prophète Mohamed

Pendant 23 ans le coran a été révélé au prophète par fragments, au gré des circonstances, apportant tel éclaircissement ou telle voie à suivre lorsque le besoin d’être guidé se fit sentir, dans cette période les accompagnons du prophète avait appris le Coran par cœur **[**[**islamweb.com 2014**](#_ENREF_22)**]**.

* 1. Le Coran à l’époque des califes (l’écriture du Coran)

La mort de plusieurs compagnons qui apprenait le Coran dans leurs cœurs lors de "Houroûb ar-ridda" (حروب الردة) a poussé le Calife Omar à proposer de regrouper le Coran dans un seul livre qui sera la référence des musulmans, cette mission a été confiée au compagnon Zayd ibn thâbit qui a donné les fruits de son travail un an plus tard. **[**[**fleurislam.com 2014**](#_ENREF_18)**]**

Le territoire de l’islam s’est étendu jusque sur l’Afrique du nord et l’Asie, Les convertis de chaque région ont appris des passages du Coran auprès du ou des compagnons installés chez eux. Or il existe des variantes entre les prononciations des différents compagnons, ce problème a poussé le calife Othmâne à reproduire plusieurs exemplaires de la copie gardée par Hafsa en y intégrant les différentes variantes de lecture. **[**[**fleurislam.com 2014**](#_ENREF_18)**]**

* 1. L’édition du Coran

Actuellement le Coran est édité en milliard de copie, on trouve le Complexe du Roi Fahd pour l'impression du Saint Coran, c’est une entreprise d’édition, il édite le saint Coran en arabe et traduit vers plusieurs langues.

En 2000 le complexe a imprimé 138 million copie dont la majorité est distribué gratuitement.

**[**[**kingfahdbinabdulaziz.com 2014**](#_ENREF_26)**] [**[**CHELLI 2010**](#_ENREF_16)**]**

* 1. L’informatisation du Coran

L’utilisation de la langue arabe en informatique a commencé depuis les années soixante du 20ème siècle, au début des années 80 on a marqué l’apparition de la première application qui traite le Coran. **[**[**خضر 2004**](#_ENREF_52)**]**

Avec le développement de l’internet, le Coran était parmi les premiers sujets qui intéressent les musulmans, on trouve maintenant une dizaine de sites autour du Coran et ses sciences notamment le tafssīr et la traduction vers plusieurs langues. **[**[**خضر 2004**](#_ENREF_52)**]**

L’informatisation du coran sera traitée en détail ultérieurement dans ce chapitre.

1. Structure du Coran

Comme nous l’avons déjà défini le saint coran a été révélé par segment, il est subdivisé en plusieurs sous-ensembles, nous allons décrire la structure du Coran dans le tableau ci-dessous suivie par une description de quelque éléments

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Structure principale | les emplacements spéciaux | Structure du Mushaf | Ecriture du Mushaf | Structure incorporelle |
| la sourate | **le début de la sourate** | **la page** | **le script othmani** | **mot** |
| l’aya | **la fin de la sourate** | **le thumn** | **les al-sawāmit** | **mot-clé** |
| le mot | **fin de l’aya** | **le rub’** | **les harakāt** | **Expression** |
| La Lettre | **virgule coranique** | **le nisf** | **Al-hamza** | **unité d'objectif** |
|  | **Sajdah** | **le hizb** | **l’écriture des mots** |  |
|  | **Waqf** | **le juz’** | **les diacritiques, des signes de distinction des lettres similaires** |  |
|  |  |  | **signes phonétiques** |  |

Table 1 : Structure du Coran [ben djamaa]

* 1. Aya

C’est la plus petite unité du texte coranique, littéralement signifie un signe que l’on traduit par ‘verset’, comme le montre la figure 1, chaque aya possède un numéro dans une sourate, l’ordre des ayates a été indiqué par le prophète Mohamed. **[**[**alkahflejournal.com 2014**](#_ENREF_2)**]**

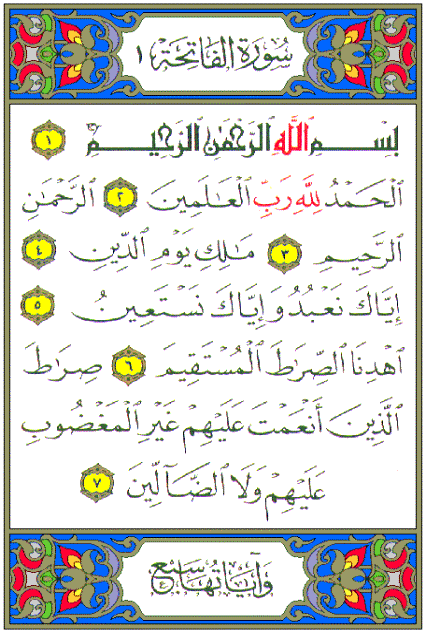


Figure 1 : Les versets de sourate al baqara

Il existe deux types de verset

* Certains sont muhkam (محكم) : avec un sens simple, claire et évident
* D’autres sont mutashabeh (متشابه) : avec de multiples interprétations possibles, alors qu’une seule qui est acceptable.
  1. **Sourate**

Le Coran est constitué de 114 sourates, chaque sourate commence par l’entête basmala sauf sourate al tawba qui ne comporte pas de basmala.

Il existe plusieurs caractéristiques pour les sourates, chacune possède :

Un nom unique, nombre de versets, un Ordre dans le mushaf, un Ordre de révélation, et un type de révélation (Mecquoise ou la Médinoise).

Les sourates du saint Coran sont classifiées en fonction de la période de révélation

Post ou pré-hégire

* **Sourates Mecquoises (Pré-hégire) :**

Ce sont les sourates qui ont été révélées avant la migration même si c’était en dehors de Mecque, ces sourates traitent le fondement de la religion (اصول الدين), ainsi que les contes sur les prophètes et les communies précédentes.

* **Sourates Médinoises (Post-hégire)**

Ce sont les sourates qui ont été révélé après la migration même si c’était en dehors de la Médine, elles parlent des hypocrites (El Mounafikine), les dialogues avec les juifs et les chrétiens, et fixent le domaine du licite et de l’illicite.

Il existe une autre classification selon le nombre de versets, les sourates sont répartis en 4 classes **[**[**alkahflejournal.com 2014**](#_ENREF_2)**]**:

* **Al tiwal** (الطوال) :

Ce sont les sept longues sourates entre Al-Baqarah (البقرة) et At-Tawbah (التوبة) [1 à 9].

* **Al mi’oun** (المئون) :

ses sourates comportent un peu plus de cent versets, entre sourate Yunus (يونس) à sourate Fatir (فاطر) [10 à 35] .

* **Al mathani** (المثاني) :

ses sourates comportent moins de cent versets, entre sourate Ya-sin (ياسين) et Al-Hujurat (الحجرات) [36 à 49]

* **Al moufassal** (المفصل) :

Elle est constituée par les courtes sourates entre Qaf (ق) et An-Nas (الناس) [50 à 114].

* 1. **Fragmentation dans le mushaf**

Afin de faciliter et organiser l’apprentissage et la lecture du saint Coran les savants musulmans se sont intéressés à la fragmentation du mushaf

Le premier qui a fragmenté le Coran était [Al-Hajjaj ben Yusef](http://fr.wikipedia.org/wiki/Al-Hajjaj_ben_Yusef) (الحجاج بن يوسف) qui voulait lire le Coran dans sept jours, donc il l’a répartie en sept fragments en se basant sur le nombre de lettres.

D’autres ont voulu le fragmenté en 30 fragments appelé le juz’ (جزء) pour pouvoir lire le Coran dans un mois, le juz’ à son a été décomposé en hizb (حزب), nisf (نصف) et rub’ (ربع), comme l’illustre la figure ci-dessous.

Figure 2 : Fragmentation d'un juz' du Coran

Il existe d’autre fragmentation adopté par quelque pays notamment la Fragmentation en manāzil (منازل) Au Pakistan et la Fragmentation en ruku‘s (ركوعات) aux pays de semi continent indien

1. L’informatisation du Coran

Le Coran possède un statut sacré chez les musulmans, la croissance des sciences coranique et des études centrées sur le saint coran ont rendu difficile aux gens de la science de faire leurs recherches, cependant, cette croissance est suivie par l’apparition de plusieurs technologies qui facilitent le stockage, la recherche et l’accès à l’information.

L’un des objectifs principaux de l’informatique est d’améliorer le temps d’accès, la recherche, la précision et la flexibilité, pour le coran et avec la quantité d’informations et des sciences s'y rattachant, l’informatisation du Coran facilite l’apprentissage, la récitation, le tafssīr, traduction, ainsi que la recherche.

Il y a plusieurs facteurs qui ont participé à l’évolution de l’informatisation du Coran et ses sciences notamment :

* L’intérêt accordé par les musulmans aux coran et sciences coraniques.
* La compatibilité de la majorité des systèmes d’exploitation actuels avec la langue arabe.
* Les symboles et les diacritiques qui sont supporté par l’encodage Unicode.
* L’apparition des font pour écrire le scripte Othmani.
* Le passage vers le développement sous des licences open source.
* L’existence des ressources Coraniques fiables et des API facilitant le développement
  1. Projets et Technologies de l’informatisation

L’apparition de certaines technologies de l’informatique et le développement des autres à entraîner une augmentation de nombre de projets lié à l’informatisation du Coran, ci-dessous quelques technologies et travaux sur l’informatisation du coran

* **Numérisation du Coran (cas de projet Tanzil)**

Un projet coranique vise à fournir un texte Coran précis et vérifié. La première version du texte Tanzil Coran est sortie en Janvier 2008 **[**[**tanzil.net 2014**](#_ENREF_44)**]** .

Ce portail web offre plusieurs fonctionnalités, on trouve un Mushaf électronique, plusieurs récitations, la traduction vers plusieurs langues, la recherche, tout en assurant une comptabilité avec les scriptes utilisés dans les mushaf réels.

De plus le portail fournie plusieurs format de texte Coranique notamment en XML, TXT et un script MySQL pour une base de données du coran.

* **Les ontologies**

L’ontologie représente la connaissance sous forme de concepts et les relations entre ces concepts, nous trouvons aussi pas mal de projets et travaux lié à l’ontologie du saint Coran, en particulier le projet « l’ontologie des concepts coranique », c’est un projet de l’université de Leeds fait par Kais Dukes **[**[**CHELLI 2013**](#_ENREF_15)**].**



Figure 3 : vision sur Qur’an Concepts Ontology

Les ontologies améliorent la qualité de recherche dans le coran en prenant en considération le contenu sémantique des mots recherchés.

* **Technologies mobiles et web**

Après la révolution des technologies mobiles, on trouve maintenant une dizaine d’application Coraniques sur les différentes plateformes mobiles, ce qui mène à plus d’innovation à savoir la reconnaissance vocale et l’amélioration des applications web existantes et plus d’interaction avec les dispositifs mobiles en se basant sur les web services exemple **Alfanous**

* 1. Problèmes de l’informatisation du Coran

Il est clair que l’informatisation du saint Coran a pris une place particulière dans les projets qui intéressent les informaticiens et les gens spécialisés dans les sciences coraniques, en parallèle il y a plusieurs problèmes qui empêchent l’évolution de l’informatisation du coran, parmi ces obstacles nous trouvons :

* **Problèmes lié à l’encodage et polices de la Langue Arabe et le script Othmani**

Pour afficher le texte coranique en Unicode avec le script othmani, il faut utiliser des

Polices spéciales, dans la police “Msh Quraan” les codes utilisé pour le symbole qui marque la fin des ayas, représente les parenthèses dans d’autre polices ce qui pose un problème de compatibilité, chose qui peut être résolue avec une police destiné aux textes coraniques **[**[**CHELLI 2010**](#_ENREF_16)**]**.

Plusieurs projet sont mise en place pour remédier à ce problème, Quran complex a développé et amélioré plusieurs polices utilisé dans l’édition du texte coranique.

En 2011 une police a été lancée par le complexe vérifié et validé par des spécialistes dans les sciences coraniques et compatible avec le codage Unicode.

* **Fiabilité des sources numérique du coran**

Les logiciels autour du saint Coran augmentent de plus en plus, mais ça ne prouve pas sa fiabilité à 100 %, nous trouvons l’exemple du logiciel Galoon, exactement dans l’aya 38 de sourate d’Al-Qalam **[**[**CHELLI 2010**](#_ENREF_16)**]**.

Pour remédier à ce problème plusieurs versions de texte coraniques authentique sont mises en ligne sous des formats standards pour faciliter leur utilisation dans différentes plateformes de développement.

* **Standardisation des références Coraniques**

Le nombre croissant des projets de numérisation du saint coran pose un problème de standardisation, dans plusieurs travaux, les développeurs ne se basent pas sur des travaux déjà réalisé, ce qui provoque une perte de temps voir de fiabilité.

Un autre problème  lié à la standardisation, en effet, on trouve  dans la traduction du Coran, qu'il existe plusieurs traductions pour la même langue.  
Plusieurs initiatives individuelles pour standardiser les références coraniques exemple ben djamaa (بن جماعة) qui a proposé une approche pour réaliser une encyclopédie standard pour le coran et ses sciences.

1. Conclusion

En raison de l’importance du coran chez les musulmans et les non musulmans et la quantité riche d’informations qu’il contient ainsi que le nombre de sciences Coraniques, le monde passe obligatoirement vers l’informatisation du Coran,

Nous allons nous baser dans notre projet sur les travaux déjà réalisés et qui ont montré leur fiabilité, toute en évitant les obstacles posé par l’informatisation du Coran.

Dans le chapitre suivant nous allons parler des éditeurs de textes, leur historique, leurs fonctionnalités, et de quelques projets similaires à notre éditeur de textes Coraniques.

## Chapitre II : Les éditeurs de texte

1. Introduction

L’histoire des éditeurs de texte remonte un peu loin, en effet le terme « traitement de texte » a été inventé par IBM à la fin des années 1960. A cette époque-là, les systèmes d’édition de texte nécessitaient des supports matériels pour fonctionner (papier, bobine,…).

Mais pas plus d’une dizaine d’années après cela, avec le besoin essentiel du traitement de texte dans divers domaines et la nécessité d’automatiser cette opération, plusieurs sociétés se sont lancées dans cette quête donnant ainsi un large choix au public, et rapidement une véritable guerre technologique et stratégique éclata entre ces firmes. Ceci a permis de réaliser nombre d’innovations dans le domaine du traitement de texte allant des éditeurs de texte grand public, éditeurs de texte scientifiques jusqu’aux éditeurs en ligne avec des outils de gestion des versions entre plusieurs utilisateurs d’un même document.

Dans ce chapitre nous allons commencer par présenter un bref historique des éditeurs de texte, puis citer les principales fonctionnalités des éditeurs de texte, en détaillant d’avantage la partie de l’écriture arabe, son encodage ainsi que les signes d’annotation coraniques.

1. Définitions
   1. Les éditeurs de texte

Un éditeur de texte est un programme informatique qui permet à un utilisateur de saisir, modifier, enregistrer, et généralement d'imprimer du texte (chiffres lettres et caractères spéciaux,). Typiquement, un éditeur de texte fournit un écran d'affichage (une page déroulante) avec une longueur fixe et les numéros de ligne visibles. Il est possible alors de remplir les lignes avec du texte, ligne par ligne. Des commandes spéciales permettent de sauter une ligne, d’insérer un espace, de déplacer vers une nouvelle page, d’effectuer des modifications globales dans le document, d’enregistrez le document, et d'effectuer d'autres actions. Après avoir enregistré un document, il est possible ensuite de l'imprimer ou de l'afficher. Avant de l'imprimer ou de l'afficher, il est possible de le formater pour certains types de fichiers spécifiques.

* 1. Encodage de caractères

Un jeu de caractères codés est un code qui associe un jeu de caractères abstraits d’un ou plusieurs systèmes d’écriture (comme des alphabets ou des syllabaires) utilisés pour transcrire des langues naturelles avec une représentation numérique pour chaque caractère de ce jeu, ce nombre pouvant lui-même avoir des représentations numériques différentes. Par exemple, le code Morse (qui associe l’alphabet latin à une série de pressions longues et de pressions courtes sur le manipulateur morse du télégraphe) et le code ASCI (qui code les lettres, les chiffres et d'autres symboles comme des entiers codés sur 7 bits) sont des jeux de caractères codés.

* 1. Polices d’écriture :

C’est un ensemble de caractères partageant le même style, Pour permettre l’écrire du texte avec différents styles ou présentations. Parmi les polices les plus connues on peut trouver Arial, Calibri, Times New Roman **[**[**youtips.com 2014**](#_ENREF_49)**]** ,nous distinguons deux types principaux de polices :

* **Les polices matricielles** : dans ce type de police les caractères sont représenté par des matrices binaires (images qui généralement sous format bitmap), le problème de ce type c’est qu’elle ne s’adapte pas au changement de la résolution.
* **Les polices vectorielles** : Ce type complète les limites des polices matricielles, car elles effectuent une suite d’instructions graphiques pour dessiner les caractères, un caractère une juxtaposition de traits et d'arcs et non pas une image.

1. Historique

Comme cité plus haut, l’histoire des éditeurs de texte date des années 60 ou les éditeurs de texte des machines à écrire plutôt que de simples programmes utilisateurs, nous citons ici un bref aperçu historique :

* En 1960 le terme « Traitement de texte » a été inventé par IBM
* En 1969, deux produits logiciels d'édition de texte (Astrotype et Astrocomp) ont été développés et commercialisés par Information Control Systems. Les deux produits utilisaient le mini-ordinateur Digital Equipment Corporation PDP-8, des lecteurs de bande accessibles de manière aléatoire, et une version modifiée de la machine à écrire IBM Selectric.
* En 1971, le traitement de texte a été reconnu par le New York Times comme un "buzz word"
* Au début des années 1970, débute une guerre acharnée entre les différentes sociétés d’édition de textes (VI, Emacs).

1. Fonctionnalités

Les fonctionnalités des éditeurs de texte diffèrent grandement en fonction du domaine d’application, cependant, nous allons recenser les principales fonctionnalités que prennent en charge la majorité des éditeurs de texte :

* Ouverture d’un fichier (en proposant parfois une liste de fichiers récemment ouverts, ou déjà existants, voire en permettant de restreindre cette liste par un filtre)
* Ajout de texte dans une ligne, des lignes dans une page, ou des pages dans un document
* Suppression de caractères d’une ligne, des lignes d'un fichier, ou des pages d’un document
* Recherche/remplacement d’une chaîne de caractères
* Le [complètement automatique](http://fr.wikipedia.org/wiki/Compl%C3%A8tement) Intelligent (IntelliSense).
* Sauvegarde du fichier, ou au contraire sortie en renonçant aux modifications (en cas de grosse erreur comme un effacement involontaire de texte).
* Faire du [Copier-coller](http://fr.wikipedia.org/wiki/Copier-coller) / Couper-coller
* Ouverture simultanée de plusieurs fichiers
* Gestion de zones tampon entre les fichiers
* Gestion des abréviations à la volée
* Fonctionnalités avancées de recherche et remplacement de texte ou motifs (notamment par l'usage des [expressions rationnelles](http://fr.wikipedia.org/wiki/Expression_rationnelle))
* Interaction avec des programmes externes sur les fichiers ([compilateurs](http://fr.wikipedia.org/wiki/Compilateur) notamment)
* [Indentation](http://fr.wikipedia.org/wiki/Style_d%27indentation) automatique pour certaines [extensions de fichiers](http://fr.wikipedia.org/wiki/Extension_de_fichier), comme le [code source](http://fr.wikipedia.org/wiki/Code_source) de divers [langages de programmation](http://fr.wikipedia.org/wiki/Langage_de_programmation).
* [Coloration syntaxique](http://fr.wikipedia.org/wiki/Coloration_syntaxique)
* Prise en charge de plusieurs langues
* Prise en charge des options de mise en forme / mise en page du document
* Possibilité d’exporté le document sous divers format de sorties : PDF, PostScript…
* Possibilité d’écrire de droite à gauche

Ces fonctionnalités sont résumées dans la figure ci-dessous.

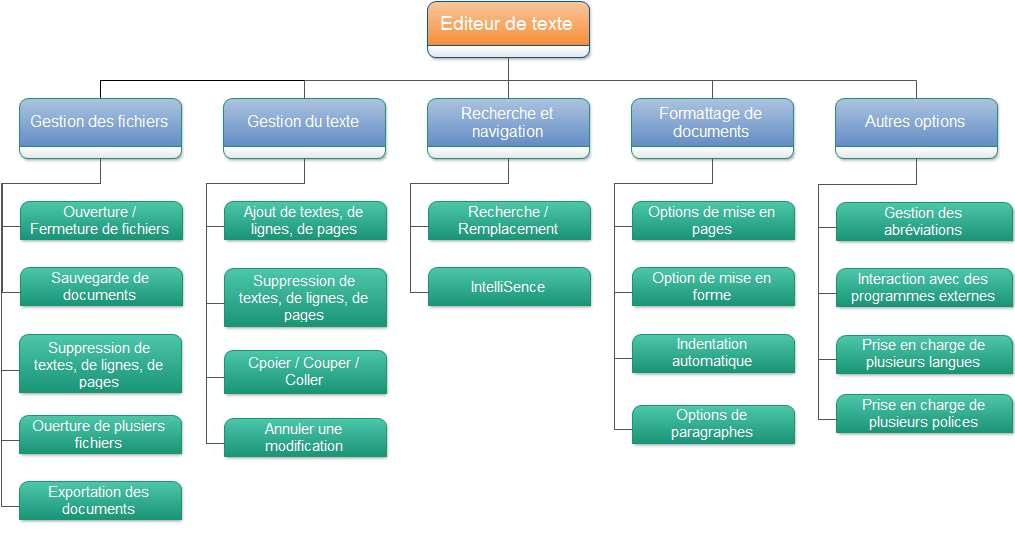


Figure : Fonctionnalités des éditeurs de texte

1. Typologie des éditeurs de texte

Nous pouvons classer les éditeurs de texte en deux grandes catégories : les éditeurs de texte basé sur le web, et les éditeurs de texte orienté desktop, ces deux catégories se ramifient tous deux en deux sous catégories qui sont : les éditeurs de texte interactifs et les éditeurs de texte automatiques.

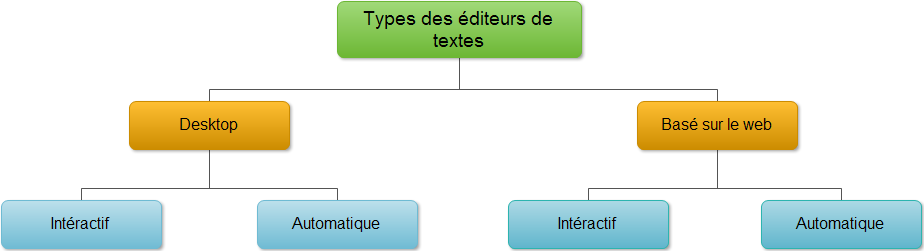


Figure 5 : Typologies des éditeurs de texte

Etant donné la des deux premières catégories, nous n’allons expliquer que les deux catégories de niveau inférieur.

* 1. Les éditeurs de texte interactifs

Cette catégorie d’éditeurs de texte propose une interface homme-machine pour permettre à l’utilisateur de contrôler la mise en forme de son document en temps réel et de manière interactive. Ce genre d’éditeurs de texte sont décrits par l’acronyme WYSIWYG « What You

See Is What You Get » qui veut dire « ce que vous voyez, c’est ce que vous obtenez ».

**Exemples** : AbiWord, Microsoft Word, NeoOffice Writer …

* 1. Les éditeurs de texte automatiques

Ce genre d’éditeurs permettent de traiter automatiquement le texte, ce sont en fait des compilateurs pour des formats de sorties bien déterminés. Il se compose généralement eux-mêmes d’un éditeur de texte pour écrire du code source, puis ce n’est qu’une fois compilé, que le texte ou le document de sortie sera généré. Ils sont généralement décrits par l’acronyme WYSIWYM « What You See Is What You Mean » qui veut dire « ce que vous voyez, c’est ce que vous voulez dire » par opposition à la précédente catégorie.

**Exemples** : Latex, Manuscript, Sprint …

1. L’écriture arabe

La langue arabe appartient à la grande famille des langues sémitiques, son écriture va de droite à gauche et les lettres de son alphabet y sont liées selon des règles se ligatures bien établies et qui sont identiques pour les deux modes manuscrit ou imprimé. Son alphabet composé de 28 lettres, 25 d’entre elles sont des consonnes, et 3 (أ و ي) sont des consonnes ou des voyelles selon leurs positions dans les termes.

La graphie des lettres diffère selon leurs apparitions dans les mots. Ainsi la lettre ع est transcrite عاد en début de mot, يعود en milieu de mot, et en fin مع isolé en fin de mot. Il en résulte 78 formes graphiques à partir de 28 lettres. Par ailleurs, la distinction minuscules / majuscules n’existe pas. **[**[**Baloul 2003**](#_ENREF_7)**]**

* 1. Les diacritiques

Les voyelles brèves sont figurées par des symboles appelés « signes diacritiques » (علامات التشكيل). Ces symboles sont absents à l’écrit dans la majorité des textes arabes ce qui peut engendrer des ambiguïtés de prononciation et de compréhension. L’opération qui consiste à insérer ces signes est appelée vocalisation automatique ou voyellation automatique. Au nombre de trois, ces symboles sont transcrits de la manière suivante :

* La *fetha* [a] est symbolisée par un petit trait au-dessus de la consonne ( بَ/ba/)
* La *domma* [u] est symbolisée par un petit trait au-dessus de la consonne ( بَ/ba/)
* La *kasra* [i] est symbolisée par un petit trait en-dessous de la consonne ( بِ/ba/)
* Le *soukoun* est symbolisée par un petit rond au-dessus de la consonne lorsque celle-ci n’est liée à aucune voyelle ( بْ/ba/) **[**[**Baloul 2003**](#_ENREF_7)**]**
  1. Le tanwine

Le signe du *tanwin* est ajouté à la fin des mots indéterminés (نكرة). Il est en relation d’exclusion avec l’article de détermination (ال) placé en début de mot. Les symboles du *tanwin* sont aux nombre de trois et sont constitués par dédoublement des signes diacritiques présentés ci-dessus, ce qui se traduit par l’ajout du phénomène /n/ au niveau phonétique.**[**[**Baloul 2003**](#_ENREF_7)**]**

* ً [an] : (باً/ban/)
* ٌ [un] : (بٌ/ban/)
* ٍ [in] : (بٍ/bin/)

On remarque l’apparition de la lettre ا [a] pour le premier *tanwin*, c’est une règle à respecter.

* 1. La chadda

Le signe de la *chedda* peut être placé au-dessus de toutes les consonnes en position non initiale. La consonne qui la reçoit est alors analysée en une séquence de deux consonnes identiques :

ّ : كلّمَ/kallama/ **[**[**Baloul 2003**](#_ENREF_7)**]**

Voyelles brèves Chedda Soukoun

وَ لَحْمِ طَيْرٍ مِمَّا يَشْتَهُونَ

Voyelles longues Tanwin

1. Codage de l’écriture arabe

En 1980, le premier code arabe standard qui est apparu est le CUDAR-U (sur 7 bit).En 1982 l’ASMO (Arab Standard and Metrology Organization) a créé le standard ASMO-449 (7 bit) qui est devenu la base des standards suivants. En 1986, SMO a standardisé un code sur 8 bits (ASMO 708), qui est devenu l’ISO-8859-6. Microsoft n’a pas opté pour le standard, elle a créé son propre codage sur 8 bits pour l’arabe pour permettre l’utilisation simultané de l’arabe et du français.

La multitude des codages rend l'échange des données multilingues difficile et parfois impossible d'où le besoin d'un codage universel.

Le codage universel Unicode (sur 16 bits) couvre la plupart des caractères des langues vivantes dans le monde entier. Le codage des caractères arabe dans l’Unicode coïncide avec le codage ASMO-449. Ce codage offre toutes les fonctionnalités du script arabe que nous allons citer ci-dessous. **[**[**ZERROUKI 2005**](#_ENREF_50)**]**

* 1. Fonctionnalités du script arabe

Nous allons citer les fonctionnalités du script arabe standard ainsi que des scripts coraniques et leurs caractères spéciaux qui sont pris en charge par le codage Unicode.

* L’ensemble des lettres est bien définie
* Les signes de ponctuations arabes : ، ؛
* Modes Liants et antis liants qui deux modes de formatage de caractères, l’un pour lier les lettres (U+200D LIANT SANS CHASSE), et l’autre pour interdire la liaison (U+200C ANTILIANT SANS CHASSE). Exemple :
  + Standard ه + ل = هل
  + Liant ه + + ل = هل ( liant)
  + Anti liant ه + = هـ ( anti liant)
* Signes diacritiques évoqués précédemment
* Chiffres arabes, arabo-hindi, arabo-hindi orientaux
* Ligatures : certaines d’entre elles sont obligatoire quelle que soit la police utilisée, exemple : لا
* Les marques coraniques du mushaf.
  1. Les signes d’annotation coranique disponibles dans l’Unicode

Le saint coran est écrit en langue arabe, néanmoins il utilise des annotations spécifique au coran, comme par exemple : les marques de fin des *ayat*, Les marques de prosternation et bien d’autres signes spécifiques à la science du *tadjwiid*.

1. L’édition de textes Coraniques
   1. Editeurs de texte coranique

Jusqu’à présent aucun éditeur de texte spécialisé dans le traitement de textes coraniques n’a été développé ou proposé, mais il y a des travaux dans ce contexte nous citons une extension sur Microsoft office qui ajoute le menu Al-Quran in Ms Word, ce plugin est disponible sur ce lien http://quran-in-ms word.software.informer.com/download/

Ce plugin assure plusieurs fonctionnalisées d’édition de texte Coraniques nous citons

* L’insertion des versets Coraniques et les signes tels que les waqf et les sajdahs
* Recherche par mot clé, par sourah et par numéro de verset.
* Utilisation de plusieurs sources de traduction.

La figure ci-dessous

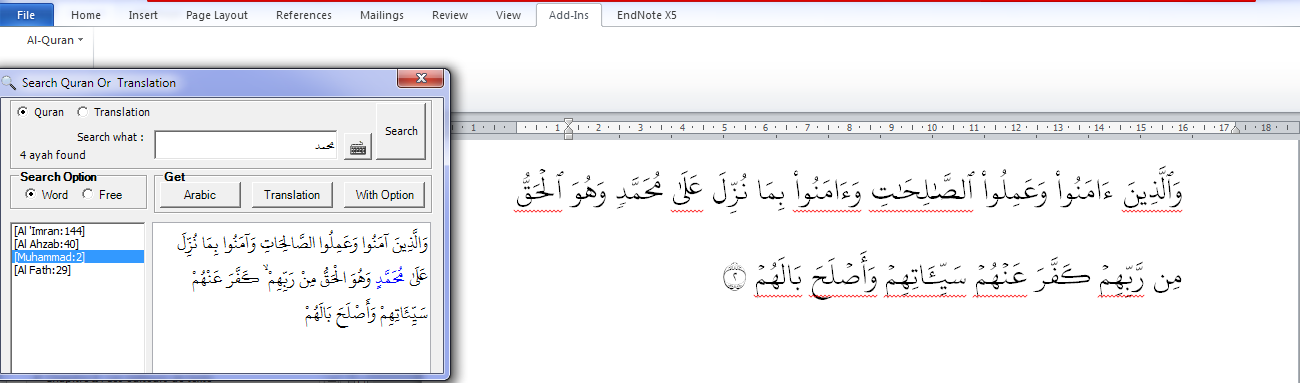


Figure 6 Insertion d'un verset dans Word en utilisant Al-Quran plugin

Cet outil facilite l’insertion des versets Coraniques mais nous pouvons citer quelque limite de ce projet notamment :

* Le plugin est mono plateforme, il n’existe que sur l’éditeur de texte Ms Word qui est un éditeur de texte payant
* L’éditeur de texte n’offre pas la recherche avancée, il se contant sur la recherche basé sur des mots clé
  1. Les polices pour l’édition de texte coraniques

Il y a plusieurs signes dans le coran qui ne sont pas supporté par les polices connue et utilisé dans l’édition des textes, ces limites ont amené à la création de plusieurs polices qui supporte la dernière version de l’Unicode nous mentionnons quelque unes

* **KFGQPC Uthman Taha Naskh**: C’est une police compatible avec l’Unicode destiné à l’écriture de texte du saint Coran selon le script Othmani dans Mushaf Almadinah.

Conçue et développé par l’équipe recherche et développement de complex roi FAHD et le calligraphe Dr Othman Taha , examiné et approuvé par le Comité scientifique d’examination du Coran **[**[**qurancomplex.gov.sa 2011**](#_ENREF_34)**]** .

يَٰٓأَيُّهَا ٱلَّذِينَ ءَامَنُوٓاْ أَوۡفُواْ بِٱلۡعُقُودِۚ أُحِلَّتۡ لَكُم بَهِيمَةُ ٱلۡأَنۡعَٰمِ إِلَّا مَا يُتۡلَىٰ عَلَيۡكُمۡ غَيۡرَ مُحِلِّي ٱلصَّيۡدِ وَأَنتُمۡ حُرُمٌۗ إِنَّ ٱللَّهَ يَحۡكُمُ مَا يُرِيدُ ١

* **ArabeysQr**: vise à soutenir pleinement la langue arabe dans l'environnement Unix / Linux. Il est conçu pour être un emplacement central pour normaliser le processus d’arabisation. Arabeyes repose sur des contributions volontaires **[**[**sourceforge.jp 2015**](#_ENREF_41)**]**

يَٰٓأَيُّهَا ٱلَّذِينَ ءَامَنُوٓاْ أَوۡفُواْ بِٱلۡعُقُودِۚ أُحِلَّتۡ لَكُم بَهِيمَةُ ٱلۡأَنۡعَٰمِ إِلَّا مَا يُتۡلَىٰ عَلَيۡكُمۡ غَيۡرَ مُحِلِّي ٱلصَّيۡدِ وَأَنتُمۡ حُرُمٌۗ إِنَّ ٱللَّهَ يَحۡكُمُ مَا يُرِيدُ ١

L’avantage principal des polices mentionnées ci –dessus c’est l’utilisation du codage Unicode, ce qui permet de conserver le texte en cas de changement de police, chose qui n’est pas assurée par les polices à symboles (exemple : Msh Quraan) cet avantage est illustré dans la figure ci-dessous

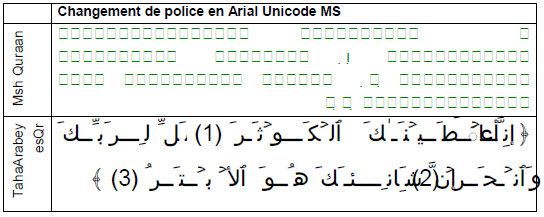


Figure 7 : Changement des polices Msh Quran et ArabeysQr en Arial unicode

1. Conclusion

Dans le chapitre suivant nous aborderons un point très important et très lié aux éditeurs de texte, ce sont les moteurs de recherches qui sont le cœur des fonctions de recherches avancées et de [complètement automatique](http://fr.wikipedia.org/wiki/Compl%C3%A8tement) Intelligent (IntelliSense) dans les systèmes de recherche

## Chapitre III : Les moteurs de recherche

1. Introduction

La quantité gigantesque d’information contenu aujourd’hui sur le net, le besoin de les extraire et les exploiter ont donné naissances a plusieurs moteurs de recherche, en particulier Google qui a bouleverser le monde de l’informatique en général et du web en particulier.

Lorsqu’on parle des moteurs de recherche, on pense directement aux fameux moteurs de recherche sur le web (Google, Yahoo, Bing), on les trouve dans la majorité des applications informatique, dans les éditeurs de texte, les systèmes d’exploitation, et les plateformes web, car ils permettent de localiser plusieurs type d’information (texte, vidéo, image) dans un temps réduit et avec plus de fonctionnalité.

Notre éditeur de recherche est composé d’un moteur de recherche de texte coranique, pour cela dans ce chapitre nous allons parler des moteurs de recherche, définir les principaux concepts lié aux moteurs de recherche, leurs fonctionnalité, leur architecture, puis nous concentrons sur les moteur de recherche Coraniques et les principales travaux similaires.

1. Définition
   1. Les moteurs de recherche

Un moteur de recherche est une application qui fournit un ensemble de réponses à une requête, le type de résultats diffère d’un moteur à autre. **[**[**AlKharashi 1999**](#_ENREF_3)**]**

Un moteur de recherche est un logiciel permettant de retrouver des ressources (pages

Web, images, vidéo, fichiers, texte ... etc.) Associées à des requêtes.

Certains sites web offrent un moteur de recherche comme principale fonctionnalité ; on

Appelle alors moteur de recherche le site lui-même (Google, Yahoo, Bing… sont des moteurs

De recherche). **[**[**NEJJARI 2007**](#_ENREF_33)**]**

Ce sont aussi des outils de recherche constitué des robots permettant de parcourir les pages web dans des intervalles réguliers de façon automatique et les indexer dans des bases de données gigantesques qui seront exploité par les internautes. **[**[**Sanan 2008**](#_ENREF_38)**]**

Il y a deux types de moteurs de recherche, classique et sémantique, la différence entre les deux ne figure pas dans leurs interface mais dans leurs architecture les premier fournies des résultats à partir d’un simple lexique, par contre les deuxième se basent sur l'idée de traitement d’informations disponibles sur le web en fonction de leurs significations. **[**[**CHELLI 2010**](#_ENREF_16)**]**

* 1. **Requête**

Une requête signifie un besoin d’information par un utilisateur, il y a plusieurs types de langages d’interrogation, Une requête peut être exprimée :

* En langage naturel ou quasi naturel **[**[**Robertson 1999**](#_ENREF_36)**]** **[**[**Salton 1971**](#_ENREF_37)**]** (exemple, "trouver toutes les usines de fabrication de voitures et leurs adresses”)
* Dans un format structuré, appelé aussi interrogation en langage booléen **[**[**Bourne 1981**](#_ENREF_9)**]**(exemple, ”voitures et marque et usines”)
* En langage graphique à partir d’une interface graphique. **[**[**Lelu 1992**](#_ENREF_28)**]**

1. Fonctionnalité d’un moteur de recherche

L’élément en commun entre tous les moteurs de recherche est la barre de recherche pour effectuer la recherche simple qui est le mode de recherche par défaut, ce mode de recherche est utilisée pour des requêtes simples, dans des cas complexe comme l’exemple d’une recherche dans des fichiers docx on fait appel à d’autres options de recherche

* 1. La recherche avancée
* **Recherche par champs (Field search)** : La majorité des documents sur le web possèdent une base de données documentaire contenant au moins quatre champs le titre, l'auteur, les sujets et la source.

La majorité des moteurs de recherche offrent la possibilité de chercher selon ses champs, l’exemple dans cette requête « intitle:Coran filetype:pdf », cet exemple demande la recherche de mot "Coran" situé dans le titre de la page, et le type fichier doit être PDF. **[**[**CHELLI 2010**](#_ENREF_16)**]**

* **Recherche par une partie de mot :** parfoisl’utilisateur commet des erreurs d’orthographe, la recherche par partie de mot peut donner les résultats souhaités par l’utilisateur, la figure ci-dessous un exemple de recherche par partie de mot

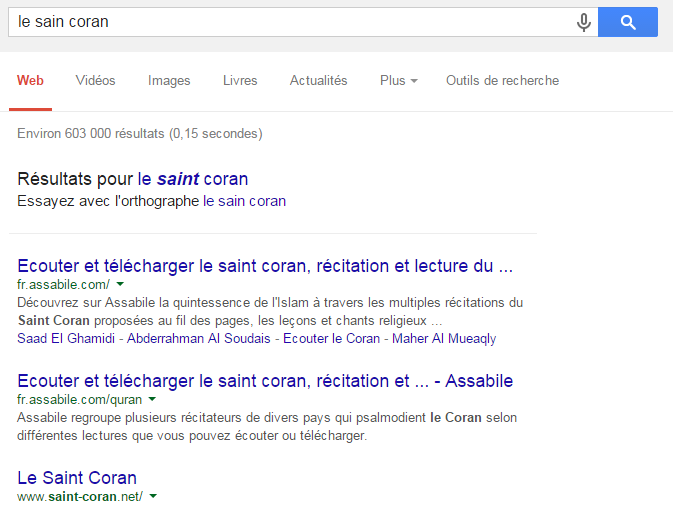


Figure 4 : La recherche par partie de mot

* **Recherche par les jokers (les caractères spéciaux) :** on utilise les expressions génériquesLorsqu’on veut supprimer une chaine de caractères dont on veut ignorer la valeur.

Si nous voulons par exemple Quran ou Kuran nous pouvons utiliser ?uran

* **Recherche par une phrase :**

La recherche par phrase permet de restreindre le nombre de résultat de la requête elle est utile dans les cas de la recherche des titres des ouvrages, des citations, des proverbes, des noms … etc ,

* **Recherche par les relations logiques**

On peut utiliser les expressions logiques pour la recherche surtout :

**Et** : pour retourner les résultats contenant tous les éléments de la requête exemple : « Coran et informatique ».

**Ou** : pour retourner les résultats contenant au moins un élément de la requête exemple « Coran ou Quran ».

**Non :** pour retourner les résultats qui ne contiennent pas un élément de la requête exemple « Coran et non Quran ».

**() :** On peut utiliser les parenthèses pour former des requêtes complexes.

* **Recherche par proximité**

Parfois l’utilisateur veut faire une recherche entre plusieurs concepts qui doivent être proche entre eux exemple **near** et **pre**

* **Recherche par intervalle**

L’importance de cette option apparait dans le recherche des valeurs numériques, si on veut chercher des informations sur les version de SQL server de 2000 à 2008 on écrit

SQL server [2000 ,2008].

* **Boosting :** il est utilisé pour désigner l’ordre d’importance des éléments de la requête
  1. Suggestion : Dans la suggestion on trouve trois types :
* **Suggestion des requêtes de recherche :** Avec cette fonctionnalité les moteurs de recherche proposent les requêtes déjà faites par d’autres utilisateurs, la figure ci-dessous illustre cette option.

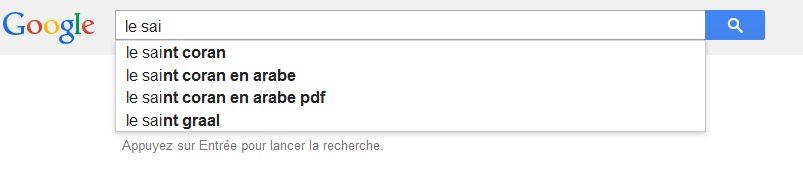


Figure 5 : La suggestion des requêtes de recherche

* **Suggestion des mots clé alternatifs**

Lorsqu’on écrit un mot inexistant dans le dictionnaire ou on commet une erreur d’orthographe le moteur de recherche propose des mots clé alternatifs, un exemple de ce type de suggestion montré par la figure ci-dessous

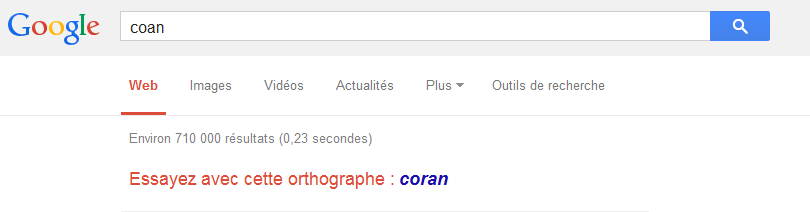


Figure 6 : Suggestion des mots clé alternatifs

* **Suggestion de mots-clés apparentés [**[**CHELLI 2010**](#_ENREF_16)**]**

Le moteur de recherche suggère des mots clé sémantiquement lié avec les mots recherché, dans la figure ci-dessous nous avons fait une recherche sue « Islam », nous avons trouvé « coran » parmi les mots clé apparenté

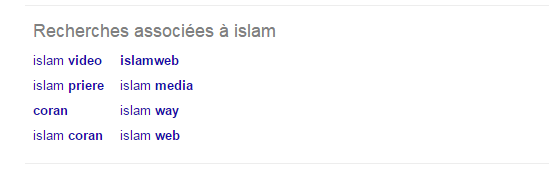


Figure 7 : Suggestion de mots-clés apparentés

* 1. Le classement

Le but que les moteurs de recherche est de fournir les résultats les plus pertinentes pour satisfaire les utilisateurs, chaque moteur de recherche utilise son propre algorithme de calcul de degré de pertinence pour mettre les résultats les plus importants en premier.

1. Fonctionnement des moteurs de recherche

Le parcours d'une requête commence bien avant la saisie, La force des moteurs de recherche réside dans leurs architectures et algorithmes mise en place pour gagner en termes de temps de réponse et de degré de pertinence, la figure ci-dessous montre les trois phases dans le fonctionnement des moteurs de recherche

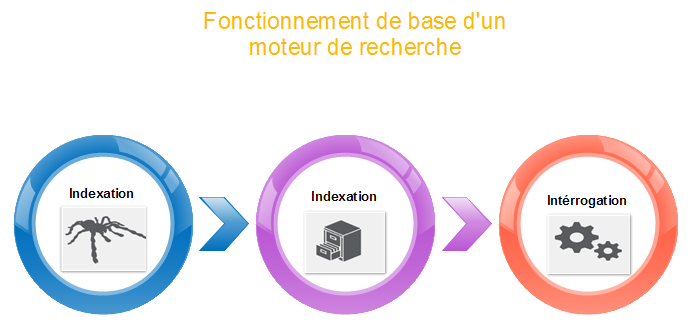


Figure 8 : Fonctionnement de base d'un moteur de recherche

* 1. L’exploration (crawling)

C’est le processus dans lequel les robots (crawlers) parcourt une quantité considérable d’informations (exemple le web) pour les indexer et soutenir un moteur de recherche, l'exploration est de recueillir rapidement et efficacement autant de pages web utiles que possible avec la structure des liens qui les interconnecte. L’explorateur est parfois nommé araignée (spider)**. [**[**Manning 2009**](#_ENREF_31)**] .**

Le tableau ci-dessous illustre les caractéristiques des explorateurs

|  |  |
| --- | --- |
| Les caractéristiques obligatoires | Les caractéristiques facultatives |
| La robustesse  La politesse | La Distributivité  L’évolutivité  L’efficacité  La nouveauté  L’extensibilité |

Tableau 2 : Caractéristiques des explorateurs

Les web crawler analyse le contenu d’une page web, le texte de la page sera indexé dans la prochaine étape, les autres liens qui se trouvent dans la page seront exploré ultérieurement.

* 1. L’indexation

Pour que la recherche soit moins couteuse en termes de temps, La seule façon pour éviter de parcourir un document pour chaque requête c’est d’indexer ce document.

* + 1. **Définition**

La norme AFNOR NF Z 47-102 1996 définit l’indexation comme «l’opération qui consiste à décrire et à caractériser un document à l’aide de représentations des concepts contenus dans ce document, c’est-à-dire à transcrire en langage documentaire les concepts après les avoir extraits du document par une analyse ». **[**[**ABAR 2009**](#_ENREF_1)**]**

L’indexation permet d’extraire les mots clés d’un document, ces mots clés seront faciles à utiliser par le moteur de recherche pendant le processus de recherche.

* + 1. **Modes d’indexation**

Il existe trois modes d’indexation selon la nature de document et les besoins d’utilisation

* **Indexation manuelle**

L’indexation manuelle est faite par des experts qui lisent l’intégralité du document pour extraire pour identifier les mots-clés qui représentent le document.

Avec le développement du traitement automatisé des documents, l’indexation manuelle parait limitée, mais elle le meilleur mode pour indexer certain type de document surtout qu’elle présente un avantage financier indéniable, l’indexation manuelle se base sur quatre points essentiels **[**[**Chartron 1989**](#_ENREF_14)**] .**

* la lecture du document dans son intégralité pour l'élaboration de l'indexation ;
* la prise en compte, dans le' choix des descripteurs, des objectifs du centre de documentation (applications) et des besoins des utilisateurs ;
* la complémentarité permanente entre les termes de l'indexation manuelle et le résumé;
* en l'absence de descripteur approprié, et lorsque l'émergence d'un nouveau concept n'est pas suffisamment explicite pour proposer un candidat descripteur, la possibilité d'utiliser un descripteur voisin ou générique
* **Indexation automatique**

L’indexation automatique utilise une méthode logicielle pour établir la liste ordonnée de tous les mots apparaissant dans les documents avec la localisation exacte de chacune de leurs occurrences et qui correspondent le mieux au contenu informationnel d'un document

**[**[**Hadj henni 2008**](#_ENREF_21)**].**

L’indexation automatique comporte les étapes suivantes **[**[**ALLAB 2008**](#_ENREF_4)**]** :

**Etape 1** : Extraction des mots simples

* Extraction des termes (tokenization) : les règles d’extraction dépendent de la langue.
* Suppression des mots vides (stop words) : ce sont des mots trop fréquents mais inutiles. Exemple : le, un, de, or, on, il…

**Etape 2** : Normalisation

* Lemmatisation (radicalisation) : par exemple inter + classe + ment, ici le radical est classe.
* Règles de transformation : suppression des terminaisons des pluriels.

L’indexation automatique permet de : **[**[**Hadj henni 2008**](#_ENREF_21)**]**

* limiter les choix parfois subjectifs de l’indexeur.
* alléger le travail requis par une indexation manuelle.
* éviter les incohérences résultant des interprétations différentes entre plusieurs indexeurs.
* réaliser une recherche exhaustive des sujets traités dans l’information analysée.
* **Indexation semi-automatique**

Dans l’indexation semi-automatique il y a une combinaison des deux méthodes d’indexation, un processus d’indexation automatique suivie par une indexation manuelle, la première permet de parcourir l’intégralité de document pour extraire la liste des mots qui le représentent, La deuxième est faite par les spécialistes pour établir les relations sémantiques entre les mots clés qui représentent le document. **[Chelli 2009].**

On peut mentionner comme un exemple de cette technique les deux logiciels : SINTEX et

ALEXDOC qui utilisent l’indexation assistée par ordinateur**. [**[**ABAR 2009**](#_ENREF_1)**] [**[**Sauvagnat 2005**](#_ENREF_39)**]**

**[**[**Hadj henni 2008**](#_ENREF_21)**]**

Le résultat de l’indexation est une représentation du document appelé index.

* + 1. **Les index :** ce sont des bases de données contenant une représentation de contenu d’un document, il y a plusieurs types d’index selon le besoin et l’utilisation technique**. [**[**CHELLI 2013**](#_ENREF_15)**]**
* **Document index**

Le document index maintien des informations sur chaque document, c’est un index avec mode d’accès séquentiel avec une largeur fixe, les entrées de l’index sont trié ordonnée par l’id du document, une entré est représenté par des données, une somme de contrôle de documents et diverses statistiques et un pointeur vers un fichier si le document a été exploré. **[**[**Brin 1998**](#_ENREF_11)**]**

* **Forward index**

Le moyen le plus efficace de construire un index, est de stocker une liste de tous les termes pour chaque document qu’on indexe, ceci est appelé forward index, la raison pour développer ce type d’index est de stocker immédiatement les mots par documents pendant que ces derniers sont parcouru.

Il n’est pas très efficace lors de l'interrogation parce que l'interrogation nécessite de parcourir toutes les entrées de l'index dans une durée spécifique pour tous les documents contenant le terme.

Le tableau ci-dessous illustre un exemple d’un forward index **[**[**Morten 2013**](#_ENREF_32)**]**

|  |  |
| --- | --- |
| Document | Terme |
| Grandma’s tomato soup | peeled, tomatoes, carrot, basil, leaves, water, salt, stir, and, boil . |
| African tomato soup | 15, large, tomatoes, baobab, leaves, water, store, in, a, cool, place. |
| Good ol’ tomato soup | tomato, garlic, water, salt, 400, gram, chicken, fillet, cook, for, 15, minutes. |

Tableau 3 : Forward index

* **Index inversé (reverted index)**

Pour mettre l’interrogation plus rapide, le plus intelligent est d’ordonner l’index par terme, ce type est appelé l’index inversé, chaque entrée associe à un terme tous les documents ou il apparait, cette opération est basé sur une transformation du forward index.

Cet index ne peut déterminer que si un mot existe dans un document particulier, il ne contient pas des informations sur l’occurrence d’apparition du terme ou de sa position dans le document, Un tel index décide quels sont les documents qui correspondent à une requête **[**[**CHELLI 2010**](#_ENREF_16)**] .**

Le tableau ci-dessous expose une partie de l’index inversé après une transformation appliqué sur l’index du tableau 3

|  |  |
| --- | --- |
| Terme | Document |
| Baobaob | African tomato soup |
| Basil | Grandma’s tomato soup |
| Leaves | African tomato soup, Grandma’s tomato soup |
| Salt | African tomato soup, Good ol’ tomato soup, |
| Tomato | African tomato soup, Good ol’ tomato soup, Grandma’s tomato soup |
| Water | African tomato soup, Good ol’ tomato soup, Grandma’s tomato soup |

Tableau 4 : Index inversé

* 1. L’interrogation

Contrairement aux deux étapes précédentes qui se font avant la recherche par l’utilisateur, l’interrogation s’exécute après la réception d’une requête par l’usager, en utilisant l’index un algorithme permet de restituer les documents qui correspondent le mieux aux mots qui constituent la requête de l’utilisateur.

À l'instar de l’indexation, le processus de l’interrogation se fait sur plusieurs étapes **[**[**CHELLI 2010**](#_ENREF_16)**]**

**Etablissement de la requête** : la requête peut être exprimée par plusieurs façons, parfois on exprime la requête sous forme de langage naturel, parfois par le biais d’une requête programmatique, on peut trouver les deux méthodes combinées dans le même moteur de recherche.

**Analyse de la requête :** dans cette étape plusieurs opérations qui ressemblent à celle de l’indexation sont appliquées sur la requête, le résultat de cette étape est un langage en commun avec l’indexation.

**Recherche de documents :** en se basant sur le langage commun entre l’indexation et la recherche cette phase lit les index et recherche les documents associés aux mots correspondants.

**Filtre et tri :** dans cette phase un algorithme calcule un score pour chaque document, ceci permet de classer les documents correspondants à la requête par ordre de pertinence, plusieurs facteurs sont pris en considération pour ce classement **[**[**Liddy 2001**](#_ENREF_29)**]**

* Fréquence d’apparition des termes
* L’emplacement des termes
* Popularité
* Date de publication de document
* Taille de document
* Proximité entre les termes de la requête
* Les noms propres

**Exposition des résultats :** Une fois la liste ordonnée des documents est prête, le moteur de recherche expose les résultats à l'utilisateur à travers une API de programmation ou dans une page web. **[**[**CHELLI 2010**](#_ENREF_16)**]**

La figure ci-dessous résume les étapes du processus d’interrogation

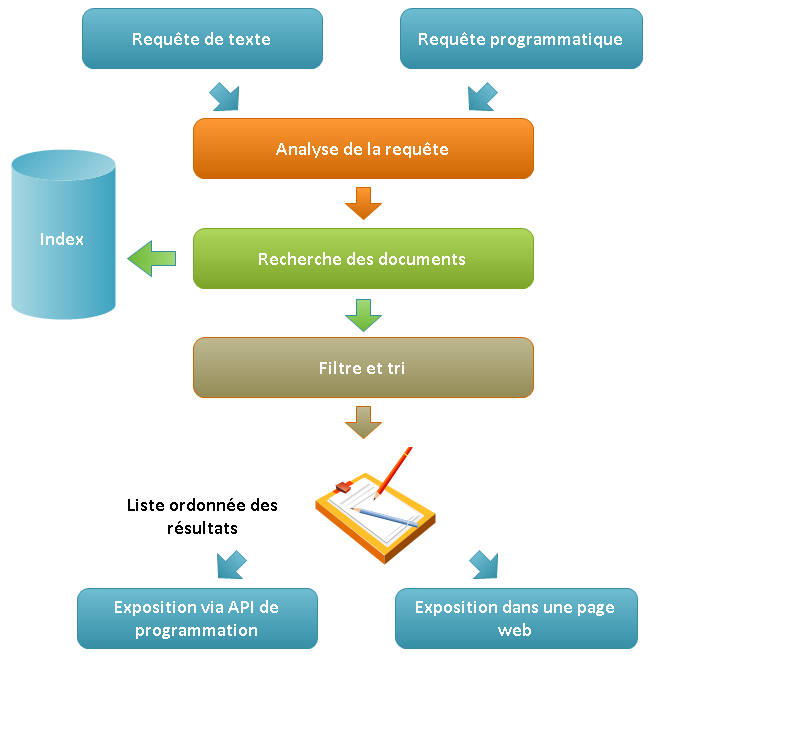


Figure 9 Processus de l'interrogation

1. Les moteurs de recherche coraniques

L’intérêt accordé au saint coran par les musulmans en général et les informaticiens en particulier à conduit à la création de plusieurs projets informatique dans la recherche dans le livre sacré

* 1. Les moteurs de recherche

Avec le développement de l’environnement web et mobile, plusieurs application de différents types permettent de répondre à des requêtes de recherche dans le coran nous citons :

* **Le projet Alfanous :** Alfanous est un moteur de recherche coranique qui offre des services de recherche simples et avancées dans l'ensemble des informations contenu dans le Saint Coran, il est basé sur l'approche moderne de la recherche d'information pour obtenir une bonne stabilité et une recherche à grande vitesse tout en assurant la majorité des fonctionnalisées offertes par les moteurs de recherche modernes .

Alfanous est disponible sur plusieurs types d’interface (desktop, mobile et web) dans différents systèmes. **[**[**alternativeto 2014**](#_ENREF_5)**]**

En plus de la recherche une API est fournie aux développeurs comme une base de plusieurs travaux qui nécessite une recherche dans le coran.

Les figures ci-dessous sont deux versions web et desktop du moteur de recherche Alfanous

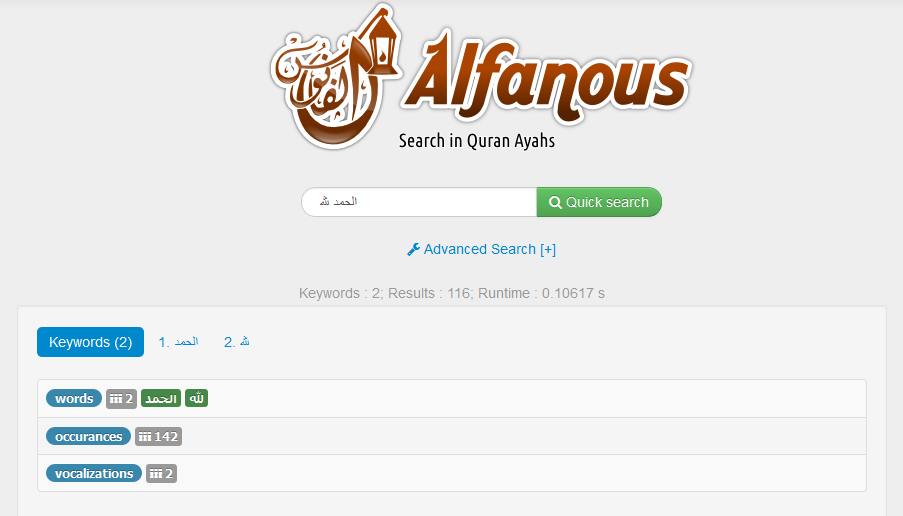
 

Figure 10 Version web Figure 11 Version desktop

* **Le projet Zekr :**Zekr est un logiciel d'étude du Coran open source pour Windows, Linux et Mac, conçu pour faciliter l'accès au texte le plus authentique et précieux des musulmans, zekr fournie plusieurs fonctionnalisées telle que Zekr fournit des traductions de Coran, la récitation, la recherche et d'autres caractéristiques à étudier le Saint Coran .

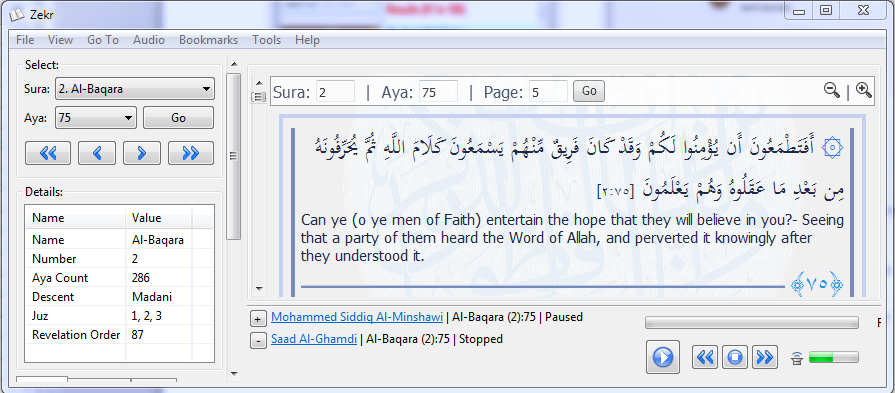


Figure 12 Interface du logiciel zekr sous Windows

* 1. Les index

Comme nous l’avons déjà cité, l’indexation est une phase importante dans chaque moteur de recherche, les index permettent d’accélérer la recherche et améliorer la qualité des réponses, nous allons parler de quelque index de Coran.

* **Les index de Zerrouki [**[**CHELLI 2010**](#_ENREF_16)**]**

C’est un effort individuel de Mr. Taha Zerrouki dans le cadre du projet de Coran et ses

Sciences, il a construit trois index qui sont aperçu dans les figures ci-dessous:

* l'index de mots
* l'index de synonymes
* l’index de sujets



Tableau 5 : Aperçu sur l’index des mots – M.Taha Zerrouki



Tableau 6 : Aperçu sur l’index des sujets – M.Taha Zerrouki



Tableau 7 : Aperçu de l’index des synonymes – M.Taha Zerrouki

* **Les index de Tanzil**

Tanzil est un projet coranique qui vise à fournir un texte précis du Coran et hautement vérifié. Il a été développé par Hamid Zarrabi-Zadeh et révisé par une équipe de vérification

du texte. **[**[**CHELLI 2010**](#_ENREF_16)**]**

En plus de texte authentique du coran Tanzil fournit également certains index comme **[**[**tanzil 2014**](#_ENREF_43)**]**

* index de la sourate.
* Index de l’aya
* index du sajdah.
* index structurel en :
* Juz’.
* Rub’.
* Manzil .
* Ruku’.
* Page.

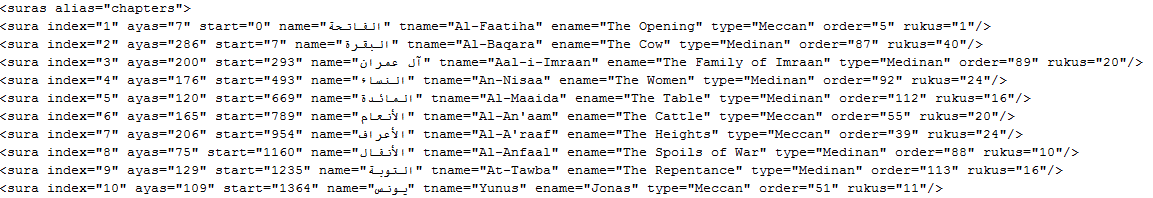


Figure 13 : Index de la sourate - Tanzil.net

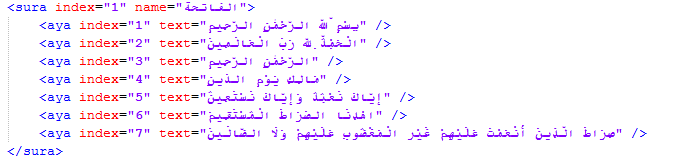


Figure 14: Index des ayates - Tanzil.net

Nous voyons d’après les exemples précédentes que chaque projet fournie des index différentes des autres, chose qui permet de combiner plusieurs index dans un seul moteur de recherche

1. Conclusion

Dans ce chapitre nous avons parlé des moteurs de recherche, leurs fonctionnalisées, leur fonctionnement, nous avons aussi cité des projets lié à la recherche d’information dans le coran qui pourront certainement nous faciliter la tâche dans l’implémentation de notre moteur de recherche une des partie importante dans notre éditeur de texte ,pour ne pas refaire des travaux déjà réalisé ,ça sera prioritaire d’exploiter des projets stables et qui ont montré leur fiabilité .

Dans le prochain chapitre, nous détaillerons le concept des web services, surtout que notre projet vise à fournir des web services exploitable par les développeurs.

## Chapitre IV : Les web services

* + - 1. Introduction

Le web contient actuellement une quantité massive de données, le nombre d’applications sur le web augmente de plus en plus, l’accès aux données pertinentes sur le web n’est pas facile car la quantité et le types de sources de données nécessite des outils évolués.

Les web services sont une des outils modernes qui combinent des anciens protocoles avec d’autres nouveaux pour permettre l’interconnexion de plusieurs applications distribué et indépendantes vis-à-vis des langages et plateformes de développement, les web services introduisent une nouvelle façon de développer des applications qui exploitent le mieux les avantages de l’internet, ils permettent de transmettre des données entre plusieurs applications.

Les web services sont caractérisés par une indépendance assurée par des protocoles de communication standardisés tels que HTTP et SMTP, et une interopérabilité très utiles surtout avec l’avancement remarquable des technologies mobiles, qui sont généralement des clients qui communique avec un fournisseur de service.

Dans ce qui suit nous allons citer les définitions des web services, leurs caractéristiques, architecture et quelques problèmes rencontrés dans l’utilisation des web services.

* + - 1. Définitions

Un service Web est un système logiciel identifié par une URI, dont les interfaces publiques et les liaisons sont définis et décrits en utilisant XML. Sa définition peut être découverte par d'autres systèmes logiciels. Ces systèmes peuvent alors interagir avec le service Web de la manière prescrite par sa définition, en utilisant des messages basé sur XML et transportés par les protocoles Internet. **[**[**Brown 2002**](#_ENREF_12)**]**

Selon **[**[**Babin 2003**](#_ENREF_6)**]** Les Web Services sont des applications qui relient des programmes, des objets, des bases de données ou des processus d'affaires à l'aide de XML et de protocoles Internet standards. Les Web Services sont des compléments aux programmes et applications existantes et servent de pont pour que ces programmes communiquent entre eux. Les Web Services ne fournissent pas d'interface usager. Ils sont utilisés afin d'envoyer des données destinées à être lues par des machines.

Les web services sont tout service qui **[**[**Cerami 2002**](#_ENREF_13)**]**:

* est disponible sur internet ou un réseau privé Intranet et accessible par une URI unique.
* Utilise le system de messagerie standard XML
* n’est pas attaché a aucun système d’exploitation ni langage de programmation
* a une auto-description via une grammaire XML commune
* est découvrable par un mécanisme de recherche simple

En résumé les web services sont des programmes accessibles via internet, ils utilisent le standard

XML indépendamment des langages ou plateforme de développement. **[**[**Cerami 2002**](#_ENREF_13)**]**

* + - 1. Architecture

Il existe deux façons de voir l’architecture d’un web service, la première par décrire des rôles de chaque acteur dans le cycle de vie d’un web service, la deuxième par examiner de la pile des protocoles utilisé dans un web service.

* 1. Architecture selon les rôles

Comme illustré dans la figure les web services se basent sur l’interaction entre trois entités principales :

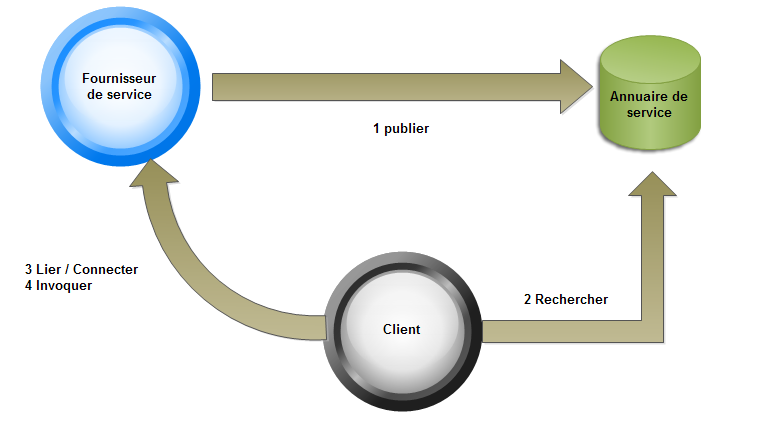
* + 1. **L’annuaire de service (Service Registry)**

C’est un répertoire de service logique et centralisé, l’annuaire fournie un emplacement dans lequel les développeurs peuvent publier des web services ou localiser les services existant, il sert donc comme une centre d'échange centralisée pour les entreprises et leurs services **[**[**Cerami 2002**](#_ENREF_13)**]**

* + 1. **Le fournisseur de service (service provider)**

 C’est celui qui implémente le service et assure sa disponibilité sur le web, d’un point de vue technique, il est constitué par la plateforme d’hébergement du service, son rôle est de définir la description de son service puis la publie dans un annuaire **[**[**Boudali 2007**](#_ENREF_8)**]**.

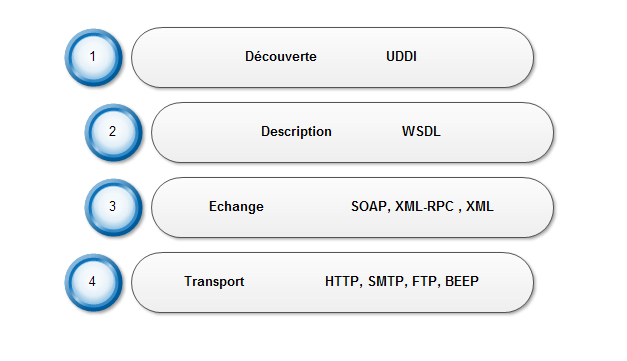
* + 1. **Le client (service requester)**

C’est tout utilisateur du web service, il consulte l’annuaire pour procéder une découverte du web service et récupère la description du web service approprié, il envoie après une requête au fournisseur sous format XML

D’après la figure plusieurs opérations se déroulent dans l’ordre suivant **[**[**Boudali 2007**](#_ENREF_8)**]** :

1. La publication du service par le fournisseur dans l’annuaire, elle est faite par la diffusion de la description du service.
2. La recherche du service qui répond aux besoins du client dans l’annuaire de services, une fois le service trouvé, sa description est récupérée par le client
3. en exploitant la description du service le client se connecte avec le serveur et pour pouvoir interagir avec l’implémentation du service considéré
4. Une fois le lien est établi entre le client et le fournisseur, le client peut désormais invoquer le service, le fournisseur à son tour répond à la requête du client
   1. Architecture selon la pile des protocoles

Plusieurs protocoles et standards sont exploités pour assurer la mise en place d’un web service, la figure ci-dessous illustre la pile des protocoles et standards utilisé dans un web service.



* + 1. **Transport**

C’est la couche responsable du transport des messages entre les trois entités participants dans le cycle de vie d’un web services, nous trouvons le protocole HTTP (Hyper text transfer protocol), SMTP (Simple mail tranfer protocol), FTP (File transfer protocol) et d’autre protocoles plus nouveaux tel que BEEP (Block extensible exchange protocol).

* + 1. **Echange**

Dans cette couche il y a l’encodage des messages dans un format XML commun compréhensible par les différentes entités, parmi les standards inclus dans cette couche SOAP, XML-RPC et XML.

* + 1. **Description**

Cette couche est responsable de la description de l’interface public du service elle indique comment communiquer avec pour utiliser le service, actuellement cette couche inclut le standard WSDL (Web service description language)

* + 1. **Découverte**

Cette couche centralise les services dans un annuaire commun et assure les fonctionnalités de publication et de recherche aujourd’hui nous trouvons dans cette couche le standard UDDI (Universal description descovery and integration)

L’architecture de base a été étendu en proposant d’autres couches tels que la couche Processus Métier et Les couches transversales (Sécurité, Administration, Transaction et QoS)

* + - 1. Les principaux standards utilisés dans les web services

Pour garantir l’interopérabilité des opérations précédentes (publication, recherche, lien et invocation), des propositions de standards ont été élaborées pour chaque type d’interactions. Nous citons, notamment les standards émergents suivants **[**[**Kellert 2003**](#_ENREF_25)**]**:

* 1. SOAP
     1. **Définition**

SOAP est un protocole léger pour l'échange d'informations dans un environnement décentralisé et distribué. Il est un protocole basé sur XML, il peut potentiellement être utilisé en combinaison avec une variété d'autres protocoles.  **[**[**Box 2000**](#_ENREF_10)**]**

SOAP assure des appels de procédures à distance RPC (Remote Procedure Call), Il comporte trois composantes principales qui sont : un framework de messagerie, un standard d’encodage et le mécanisme RPC. **[**[**Kadima 2003**](#_ENREF_24)**]**

* + 1. **Les composants principaux du protocole SOAP**

La spécification du protocole SOAP définie trois composants majeurs :

* **Le Framework de messagerie SOAP (SOAP envelop)**

Il définit des règles à respecter pour l’encapsulation des donner avant de les transporter, il inclut les données spécifiques à l’application tels que les noms des méthodes à invoquer, leurs paramètres et les types de retour. **[**[**Cerami 2002**](#_ENREF_13)**]**

Le Framwork de messagerie requiert que le message SOAP soit composé d’une enveloppe qui contient un Header et un Body. Le Header comporte les métas – données du message et le Body comporte le corps du message lui-même. **[**[**Boudali 2007**](#_ENREF_8)**]**

* **Encodage / sérialisation standard pour les objets (The SOAP encoding rules)**

Pour assurer l’échange de données entre plusieurs machines, ce dernier doivent se mettre d’accord sur les règles d’encodage des types de données spécifiques **[**[**Cerami 2002**](#_ENREF_13)**]** tels que le entiers, les chaines de caractères et les tableaux.

Le style d’encodage SOAP est basé sur la généralisation des caractéristiques communes trouvé dans les langages de programmation, les bases de données et les données semi-structurées. **[**[**Box 2000**](#_ENREF_10)**]**

* **Invocation de service d’objets distant via SOAP RPC**

SOAP permet à l’aide de XML, la réalisation d’un appel à des procédures à distance , SOAP définie des conventions simple pour représenter les appels des procédures et les réponses

* + 1. **Structure d’un message SOAP**

Chaque message échangé entre le client et le fournisseur de service doit respecter un format bien défini il doit contenir un deux éléments obligatoires un enveloppe et un corps, avec un élément optionnel qui est l’entête (voir la figure)

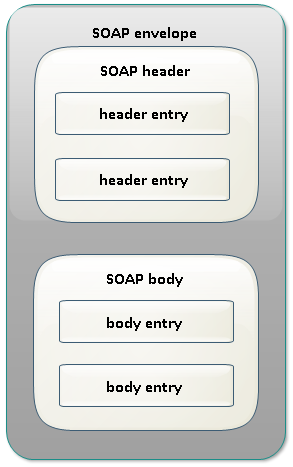


Figure 15 : Format d'un message SOAP

* **L’enveloppe (SOAP Envelope):** c’est l’élément de base d’un message SOAP, nous trouvons à l’intérieur l’espace de nom XML (XML namespace) permettant de préciser la version SOAP.

**Exemple** **[**[**Cerami 2002**](#_ENREF_13)**]**:

<SOAP-ENV : Envelope

Xmlns:SOAP-ENV="<http://schemas.xmlsoap.org/soap/envelope/>"

* **L’entête (SOAP Header) :** Cet élément optionnel offre un Framework pour spécifier d’autre information requise tel que une signature digital dans un service protégé par un mot de passe ,ou un compte de paiement du service utilisé , cet élément est facultatif et peut être utilisé dans les messages SOAP ,cependant elle assure des mécanismes d’authentification de paiement et de gestion de transaction **[**[**Cerami 2002**](#_ENREF_13)**]** **,** sonintérêt apparait lorsque le message doit être traité par plusieurs intermédiaires

**Exemple** **[**[**w3schools 2014**](#_ENREF_47)**]** :

<soap:Header>  
  <m:Trans xmlns:m="http://www.w3schools.com/transaction/"  
  soap:mustUnderstand="1">234  
  </m:Trans>  
</soap:Header>

L’exemple précédent est un message l’entête incluant une transaction (message Trans) .

Nous trouvons trois attributs principaux dans l’entête, les attributs définissent comment le récepteur doit traiter le message SOAP **[**[**w3schools 2014**](#_ENREF_47)**]**.

* **Le corps (SOAP Body) :** c’est un élément principal qui comporte des données traité par les deux extrémités le client et le fournisseur, il inclut les requêtes RPC qui contiennent le nom de la méthode invoqué par le client ainsi que les paramètres, il inclut aussi les réponses du serveur aux requêtes contenant l’information demandé par le client.

**Exemple [**[**w3schools 2014**](#_ENREF_47)**]**

<?xml version="1.0"?>  
<soap:Envelope  
xmlns:soap="http://www.w3.org/2001/12/soap-envelope"  
soap:encodingStyle="http://www.w3.org/2001/12/soap-encoding">  
  
<soap:Body>  
  <m:GetPriceResponse xmlns:m="http://www.w3schools.com/prices">  
    <m:Price>1.90</m:Price>  
  </m:GetPriceResponse>  
</soap:Body>  
  
</soap:Envelope>

<?xml version="1.0"?>  
<soap:Envelope  
xmlns:soap="http://www.w3.org/2001/12/soap-envelope"  
soap:encodingStyle="http://www.w3.org/2001/12/soap-encoding">  
  
<soap:Body>  
  <m:GetPrice xmlns:m="http://www.w3schools.com/prices">  
    <m:Item>Apples</m:Item>  
  </m:GetPrice>  
</soap:Body>  
  
</soap:Envelope>

WSDLUDDI

**Message SOAP de la réponse**

**Message SOAP de la requête**

L’exemple ci-dessus est une requête client contenant la méthode GetPrice avec le paramètre Apples (demande le prix de pommes), le serveur répond avec un message SOAP contenant le prix 1.90 dans le corps du message.

* 1. WSDL
     1. **Définition**

WSDL (Web Services Description Language) est un format XML pour décrire les services web offerts par un composant applicatif **[Kadima 2003],** WSDL décrit quatre types critiques d’information **[**[**Cerami 2002**](#_ENREF_13)**]:**

* Interface de description de toutes les fonctions publiques disponibles.
* Types de données contenues dans les requêtes et les réponses.
* La liaison avec les protocoles de transport utilisés.
* Les adresses pour la localisation des services spécifiés.
  + 1. **Spécification du WSDL**

WSDL est une grammaire basé sur XML composé de **[**[**Brown 2002**](#_ENREF_12)**] [Kadima 2003]**

**Types :** l’élément type contient les définitions des types de données appliqués aux messages échangés afin de garantir une interopérabilité maximale.

**Message :** représente une définition abstraite (noms et types) des données transmises.

**Operation :** c’est une description abstraite d’une action supportée par le web service

**Port Type** utilisé pour définir les traitement offerts par un service web ,il comporte plusieurs opérations qui peuvent être : un opération unidirectionnelle ,requête/réponse , sollicitation/réponse et notification

**Liaison (binding)**: définit le format du message et les détails du protocole pour les opérations et messages définis par un portType particulier. Il peut y avoir un certain nombre de liaisons pour un portType donné.

**Port :** définit une extrémité particulière en spécifiant une adresse unique pour une liaison.

**Service :** est utilisé pour agréger un ensemble de ports associés.

* 1. UDDI
     1. **Définition**

C’est un annuaire qui permet la publication et la découverte des web services, il peut contenir des informations sur le fournisseur et les services qu’il publie, l’inscription d’un fournisseur de service lui permet de se présenter et présenter ses services offerts **[**[**Boudali 2007**](#_ENREF_8)**]**.

Les données trouvées dans un annuaire UDDI appartiennent à trois catégories principales **[**[**Cerami 2002**](#_ENREF_13)**]**

* **Les pages blanches** : contiennent des informations générales sur le fournisseur tels que le nom, la description, adresse et numéro de téléphone ainsi qu’un identifiant unique.
* **Les pages jaunes :** ça inclut des classifications générales des données concernant le fournisseur et ses services.
* **Les pages vertes :** dans lesquels nous trouvons des informations techniques sur le web service, elles inclussent généralement des pointeurs vers des spécifications externes (WSDL) et des adresses pour invoquer le web service.
  + 1. **Le modèle de données d’UDDI (version 2.0)**

UDDI inclut un schéma XML contentant cinq structures de données fondamentales **[**[**Kadima 2003**](#_ENREF_24)**]**

* **BusinessEntity :** informations sur l’entreprise qui publie le service dans l’annuaire UDDI
* **ServiceEntity** : informations sur les services offerts par le fournisseur
* **BindingTemplate (coordonnées des services) :** informations concernant le lieu d’hébergement du service
* **tModel (descriptions techniques des services) :** informations sur le mode d’accès aux services
* **publisherAssertion :** assertions contractuelles entre partenaires dans le cadre des échanges d’exécution d’un service.
  + - 1. Conclusion

L’interopérabilité entre les applications, et l’utilisation de protocoles standards nous motiverons d’assurer des web services dans notre éditeur de texte coranique, cependant d’autres menaces vont apparaitre dans ce cas, surtout les problèmes de sécurité.

Dans le chapitres suivants nous allons parler des quelque concepts de bases de sécurité, puis des protocoles de sécurité de l’application et des web services.

## Chapitre V : La sécurité

Introduction

La sécurité est un élément essentiel dans les plateformes web, par définition les applications web permettent d’accéder à des ressources centrales à savoir des serveurs et des bases de données, la compréhension des menaces de sécurité devient primordiale avant de développer les plateformes web, la mise en place des mesures de sécurité devient primordiale pour protéger les ressources et garder un environnement sûr.

Dans ce chapitre nous allons commencer par l’introduction des concepts généraux sue la sécurité informatique, puis aborder quelque attaque et contre-mesures, par la suite nous terminons par la sécurité des web services.

Généralité sur la sécurité informatique

* 1. Définitions

**La sécurité informatique** : signifie l’ensemble des mesures permettant de protéger les données ou les systèmes permettant de traiter, stocker et véhiculer les données ,et assurer que les ressources soit utilisé uniquement dans le cadre où il est prévu qu'elles le soient. **[**[**dicodunet 2014**](#_ENREF_17)**]**

**Menace :** c’est l’action ou l’évènement dont le déclenchementpourrait porteratteinte à une ou plusieurs caractéristiques critiques de l’information et des systèmes tels que la confidentialité, l’intégrité et la disponibilité **[**[**Godart 2002**](#_ENREF_20)**]**

Nous distinguons deux types de menaces, les actes volontaires tels que le piratage, l’espionnage ,le sabotage et les actes involontaires qui sont généralement des accidents **[**[**Godart 2002**](#_ENREF_20)**]**

**Vulnérabilité** est le réactif qui permet de déclencher une menace, elle dépend du contexte **[**[**Godart 2002**](#_ENREF_20)**]** , si la menace est la destruction des données par un virus, la vulnérabilité pourra être l’absence d’un anti-virus .

* 1. Les Composants de la sécurité informatique

La sécurité est essentiellement pour protéger les caractéristiques importantes des informations et des systèmes :

* + 1. **La confidentialité :** c’est garantir que les ressources soient protégée contre toute divulgation accidentelle ou malveillante aux parties non autorisées **[**[**Godart 2002**](#_ENREF_20)**]**
    2. **L’intégrité :** L’intégrité des ressources physique ou logiques est le faite qu’elles n’ont pas été détruites ou modifiées à l'insu du propriétaire tant de manière intentionnelle qu’accidentelle **[**[**Ghernaouti 2013**](#_ENREF_19)**]**
    3. **La disponibilité :** c’est l’assurance qu’un service soit accessible et utilisable par les utilisateurs avec un temps de réponse acceptable, elle est différentes de l’accessibilité car le service doit respecter un certain temps de réponse.**[**[**Ghernaouti 2013**](#_ENREF_19)**] .**
  1. Les mesures de sécurité
     1. **Authentification :** Le service de l’authentification recouvre deux fonction, l’identification qui associe une entité physique à une entité logique représenté par un login et l’authentification pour confirmer que c’est bien l’entité physique qui se connecte, cette fonction est remplie par l’utilisation d’un mot de passe **[**[**Kadima 2003**](#_ENREF_24)**] .**
     2. **Autorisation d’accès :** son objectifs est mettre des filtres permettant e garantir que seul les flux autorisé sont accepté par ces filtres **[**[**Kadima 2003**](#_ENREF_24)**] .**
     3. **Cryptage et décryptage :** C’est l’ensemble d’algorithmes permettant de chiffrer les messages et les rendre illisibles sans une action de déchiffrement, il existe deux types de chiffrement, symétrique avec une seule clé secrète et asymétrique avec une clé public et une autre privée
     4. **Certificats :** ils sont utilisé pour établir les références d’un utilisateur ,interface ou application ,il nécessite des informations comme le nom et la signature digitale ,et la durée de validité ,les certificats nécessitent un effort de gestion pour les attribuer ,stocker et vérifier **[**[**Kadima 2003**](#_ENREF_24)**]**
     5. **Signatures :** La norme ISO 7498-2 définit la signature numérique comme des «données ajoutées à un message », ou transformation cryptographique d’un message, permettant à un destinataire de :
* authentifier l'auteur d'un document électronique
* garantir son intégrité
* Protéger contre la contrefaçon
  + 1. **Non-répudiation :** assure le fait qu’une entité ne puisse pas nier avoir effectué une activité, elle assure que le destinataire ne pourra pas nier avoir reçu l’information, et que l’émetteur ne pourra pas nier avoir émettre l’information **[**[**Godart 2002**](#_ENREF_20)**]** .

Les attaques fondamentales dans le web

Sécurité des web services

L’utilisation du XML dans les web service l’a rendu vulnérable à plusieurs attaques et menaces, souvent appelées « attaques basé sur XML » qui ne sont pas traité par les mécanismes de sécurité existants tels que les pare-feu.

* 1. Les Attaques dans les web services

Le transfert de messages dans un web service se fait par le biais du protocole SOAP et un mécanisme d’échange de données XML, ce dernier est vulnérable a plusieurs type d’attaques appartenant à deux grandes familles, comme le montre la figure suivante

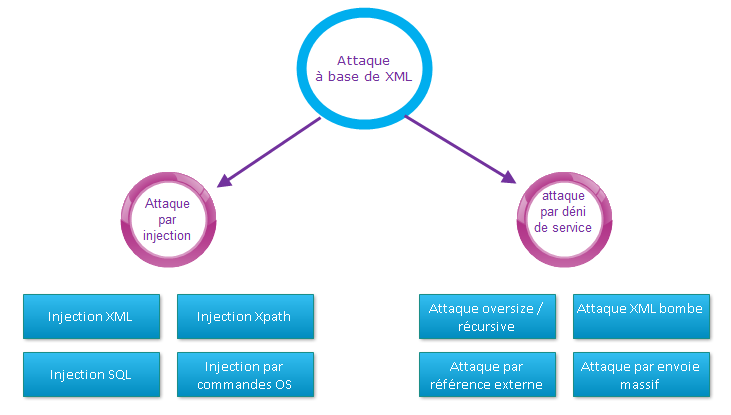


Figure 16 : Attaques à base XML

* + 1. **Attaques par injection**

Les attaques par injection sont parmi les attaques dominantes dans le web notamment les web services, les vulnérabilités qui facilitent ce genre d’attaques apparaissent lorsqu’il n’existe aucune séparation stricte entre les instructions d’un programme et les données saisies par un utilisateur **[**[**Lackey**](#_ENREF_27)**]** .

Les injections menaces les web services surtout celles qui ciblent les messages XML nous trouvons :

* **Injection XML :** elles sont utilisées pour manipuler le contenu XML, Le but de cette attaque est d’envoyer au serveur des données qui peuvent provoquer l'insertion de contenu malveillant dans le message / document résultant.

Voici les différentes formes des injections XML : **[**[**SMAHI 2014**](#_ENREF_40)**]**

* **Injection XML simple**
* **Injection de code XML malformé**
* **Injection de balises XML automatiques**
* **Injection XML persistante**
* **Injection XML dans les messages SOAP**
* **Injection Xpath :** Xpath est un langage d’interrogation des fichiers XML, il est équivalent à SQL dans le contexte des bases de données relationnelles, Mais malgré sa simplicité il pose

aussi des problèmes d’injection **[**[**Lackey**](#_ENREF_27)**]**

Lorsqu’on choisit de stocker des données sensibles en XML plutôt que dans une base de données SQL, les attaquants peuvent s’appuyer sur une injection de XPath pour contourner une authentification, comme pour inscrire des données sur le système distant **[**[**Lackey**](#_ENREF_27)**]**

Voici les différentes formes des injections Xpath **[**[**SMAHI 2014**](#_ENREF_40)**]**

* **Injection XPath Simple**
* **Dumping d’un Document XML**
* **Blind XPath Injection**
* **Injection XQuery**
* **Injection XPath dans les messages SOAP**
  + 1. **Attaques par déni de service**

XML déni de service est probablement l'attaque la plus grave contre les services Web, ce type d’attaque vise à rendre les web services indisponibles pour une durée indéterminée

Cette classe regroupe les attaques suivantes :

* **Attaque oversize :** ce type d’attaque vise l’épuisement des ressources du serveur, comme la mémoire, le processeur ou de la bande passante réseauen envoyant des messages SOAP de taille très grande ce qui engendre un déni de service pendant l'analyse du message SOAP **[**[**Jensen 2009**](#_ENREF_23)**]**
* **Attaque dite XML bombe (XML Entity Expansion)** ce type d’attaque exploite une capacité de DTD XML qui permet la création de macros personnalisées, appelées entités, qui peuvent être utilisées dans un document. En définissant de manière récursive un ensemble d'entités personnalisées en haut d'un document, un attaquant peut submerger les analyseurs qui tentent de résoudre complètement les entités en les forçant à parcourir presque indéfiniment sur ces définitions récursives.**[**[**webappsec 2014**](#_ENREF_48)**]** .
* **Attaque par référence externe :** Au lieu de définir des chaînes de remplacement d'entité en tant que constantes, il est également possible de les définir de sorte que leurs valeurs sont extraites des URI externes **[**[**Sullivan 2009**](#_ENREF_42)**]** ,l’analyseur attend une réponse de la part de la référence externe qui ne peut jamais renvoyée ,chose qui cause un déni de service .
* **Attaque par envoie massif de messages SOAP**: le but de ce genre d’attaque est de surcharger le web service en envoyant des messages SOAP valides et répétitifs, le serveur ne peut pas traiter toutes les requêtes en une période courte, ce qui empêche le serveur de répondre à des messages SOAP non malveillants **[**[**Radhamani 2007**](#_ENREF_35)**]**
  1. Mécanismes de sécurité des web services

Pour lutter contre les attaques menaçants la sécurité des web services, plusieurs mesures et standards ont été établit pour assurer la disponibilité ,l’intégrité et la confidentialité des web services

* + 1. **Sécurité XML**

Manipulation des données avec XML nécessite parfois l'intégrité, l'authentification et la confidentialité. La signature et le chiffrement XML peuvent aider à créer un environnement sûr pour XML. **[**[**w3c 2014**](#_ENREF_46)**]**

* **Signature XML** La spécification XML signature définit comment représenter des signatures digitales en XML fournissant la capacité de signer des documents entier ou des sections spécifiques **[**[**Kadima 2003**](#_ENREF_24)**]**
* **Chiffrement XML** Pour des données sensibles (financières ou personnelles) transitant sur le réseau, les expéditeurs et destinataires exigent des communications sûres. Même s'il existe déjà des technologies permettant de sécuriser l'ensemble de ces données ou la session de communication dans son ensemble, seules les signatures XML du W3C (XML Signature), utilisées en combinaison avec la nouvelle recommandation XML Encryption, permettent aux utilisateurs de signer et de crypter des parties bien définies de données XML , Par exemple, l'utilisateur d'un protocole de services Web, tel que SOAP, peut vouloir crypter seulement le corps du message XML, en laissant en clair les informations nécessaires à l'acheminement du message à son destinataire **[**[**W3C 2002**](#_ENREF_45)**]**
  + 1. **Sécurité SOAP**

Le message SOAP a besoin d’être encrypter pour protéger les contenus de changement ou d’un dévoilement inopportun .Il existe un certain nombre d’options pour atteindre ce but :

* Encrypter le corps du message SOAP et / ou les données définies dans les champs du header SOAP
* Utiliser SSL (Secure Sockets Layer) pour encrypter le trafic http.
* Etablir les connexions avec VPN qui encrypter par défaut.**[**[**Kadima 2003**](#_ENREF_24)**]**
  + 1. **Authentification et autorisation :** plusieurs spécifications ont été mise en place pourassurer l’authentification et l’autorisation dans les web services
* **SAML :** protocole couvrant les problématiques d’authentification et d’habilitation des transactions SOAP. SAML est un mécanisme bas é sur XML pour échanger des informations sur l’authentification et l’autorisation qui fournit des capacités single sign-on capabilities pour les web services **[**[**Kadima 2003**](#_ENREF_24)**]** .
* **XACML :** définit un schéma pour décrire les politiques des droits d’accèsà des objets XML, il spécifie comment exprimer des polices pour accéder à des informations sur le réseau , il gère des droits d’accès digitaux **[**[**Kadima 2003**](#_ENREF_24)**]**

Conclusion

# Partie 2 : non état de l’art

1. Introduction

Après avoir faire une étude bibliographique sur les concepts lié à l’édition du texte coranique authentique, nous avons maintenant une vision global les ces concepts, et les différents travaux déjà réalisé dans le domaine, cette partie abordera notre contribution dans la réalisation d’un éditeur de texte Coranique, aucun éditeur spécialisé dans le traitement de textes coraniques n’a été réalisé ou proposé.

Notre travail sera basé dans un premier temps sur la forme des éditeurs de texte standards afin de bénéficier des fonctionnalités déjà réalisé, après nous passerons à la personnalisation de l’éditeur de base en ajoutant des fonctionnalités spécifié aux textes Coraniques.

Cette partie nous allons présenter Noor research center l’organisme ayant proposé ce projet avec un rappel de la problématique et notre solution proposé.

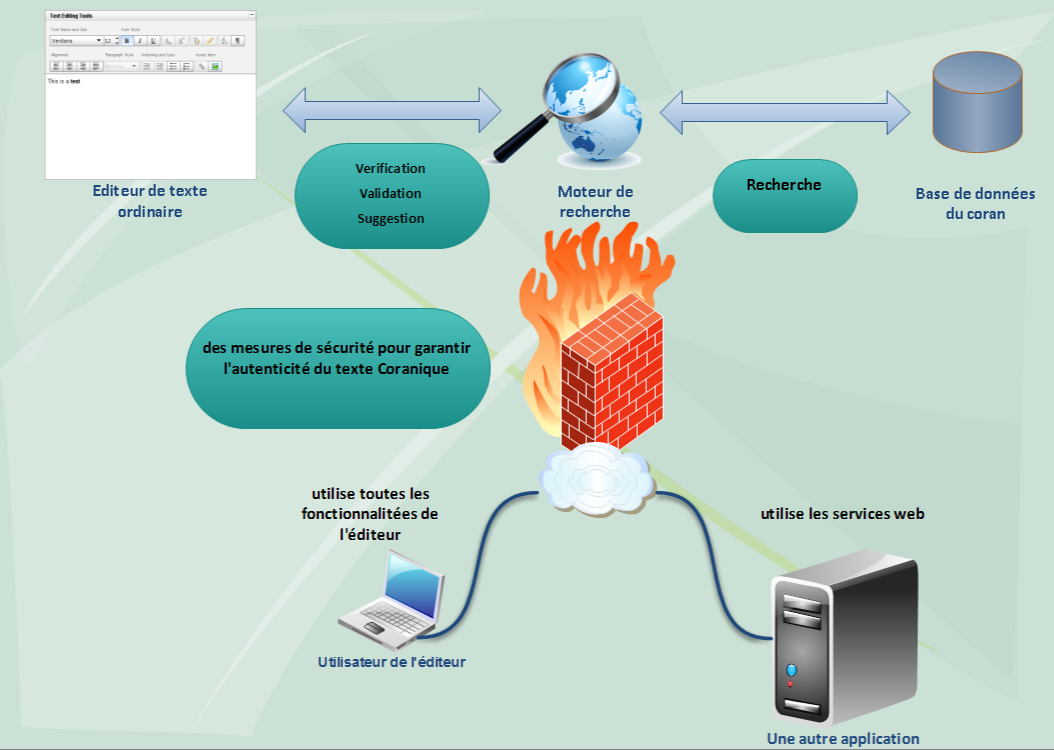
Nous allons par la suite utiliser des schémas dans la conception pour illustrer les composants de notre solution et leurs comportements, mais sans détailler comment leur implémentation, ou les outils utilisés, chose qui sera abordé dans la prochaine partie.

Les composants principaux que nous essayons d’intégrer dans notre l’éditeur de texte seront :

* Editeur de texte Coraniques basé sur des éditeurs de texte ordinaire avec l’ajout des fonctionnalités lié à la langue arabe, le Coran et le script othmani.
* Moteur de recherche Coranique faisant la collecte de données fiables et la création des index du Coran qui ne sont pas disponible et le traitement des requêtes clients (interrogation)
* Gestionnaire des web services permettant de aux utilisateurs d’utiliser les fonctionnalités offertes par notre éditeur dans d’autre applications.
* Gestionnaire des utilisateurs qui permettra de restreindre les fonctionnalités en fonction de type de l’utilisateur.
* Sécurité de l’application, l’authenticité des textes Coraniques est une partie majeure dans notre éditeur.

1. La solution proposée
   1. Schéma globale de la solution

Pour porter une solution aux problèmes posés par l’édition des textes coraniques, nous avons proposé la solution décrites ci-dessous



Notre éditeur de texte est constitué de cinq composants majeurs

* **Un éditeur de texte :**  il est basé sur un éditeur de texte ordinaire qui permet de faire l’insertion, la suppression, la modification, la mise en forme la recherche et la correction ,ces fonctionnalités sont appliqué sur n’importe quel type de texte, nous allons l’étendre pour qu’il supporte des tâches spécifié au texte Coranique tels que l’insertion des différents signes et scriptes lié au textes Coraniques , la détection , la suggestion ,la vérification de l’authenticité et la validation des versets coraniques .
* **Un moteur de recherche :** pour permettre l’insertion d’un texte coranique authentique et valide pour cela nous allons utiliser et développer des ressources Coraniques fiables approuvé par les spécialiste dans le domaine des sciences Coraniques, l’accès à ces ressources sera garantie par un moteur de recherche Coranique, il permet de localiser l’emplacement des versets demandé par l’utilisateur dans le Coran, la requête de l’utilisateur et la réponse peuvent être sous plusieurs type.

Le moteur de recherche sera utilisé par l’éditeur de texte pour rechercher dans la base de données, la vérification de l’authenticité de texte par rapport aux ressources centralisé, et la suggestion des versets selon les premiers mots inséré par l’utilisateur.

* **Un gestionnaire des web service :** ce composant permet de mettre en place une architecture orienté service pour que d’autres applications puissent, par exemple : dans une application qui traite des textes religieux, on peut utiliser les web service pour vérifier l’authenticité des versets contenue dans le texte.

Ce composant permet de mettre en place tous les protocoles qui facilitent l’accès aux services de l’éditeur.

* **Une couche de sécurité :** L’un des objectifs principaux de notre application est de fournir un texte Coranique authentique, pour cela la sécurité des ressources est très importante notamment les bases de données, la disponibilité des services, et l’intégrité des fichiers et textes transféré via le web.
* **Un gestionnaire des utilisateurs :** ce module permet de gérer les comptes utilisateurs, les droits d’accès, et les ressources de chaque compte utilisateurs.
  1. Expression des besoins

Notre éditeur de texte est composé de cinq modules, avec plusieurs types d’utilisateurs, nous détaillons les fonctionnalités de système avec les spécifications fonctionnelles et les diagrammes de cas d’utilisation.

* + 1. **Les spécifications**

Pour porter une solution aux problèmes posés par le traitement de texte Coraniques nous avons opté une liste de spécifications fonctionnelles, l’éditeur de texte Coranique que nous allons développer doit assurer les fonctionnalités suivante :

* Permettre à l’utilisateur d’éditer un texte en assurant les fonctionnalités de base de traitement de texte (ajouter, supprimer, copier, coller, police, taille, style, direction et insertion).
* Permettre d’importer des fichiers avec les formats .doc et .txt.
* Permettre de générer le texte de l’éditeur en plusieurs formats (texte, PDF, Word, image et web)
* Permettre d’insérer des versets Coraniques, leur numéro, leur Soura. Avec la prise en considération le script Othmani et tous les signes Coraniques spécifiques.
* Empêcher de modifier les versets Coraniques
* Permettre d’ajouter le tafssīr et la traduction des versets.
* Permettre d’utiliser un clavier dédié pour faciliter l’insertion du texte arabe et Coranique.
* Permettre de rechercher dans le Coran par : Numéro de verset, sujet, mots-clés, partie de mot, et par requête.
* Permettre la suggestion intelligente des versets selon les mots insérés.
* Permettre de créer un compte utilisateur, et accéder à un espace de travail
* Permettre à l’utilisateur de créer, éditer et enregistrer des documents dans son espace de travail, éventuellement dans le Cloud.
* Permettre à l’utilisateur de contacter l’administrateur.
* Permettre les partages sur les réseaux sociaux.
* Permettre aux utilisateurs de visualiser les services web disponible.
* Permettre d’utiliser les fonctionnalisées de base de l’éditeur de texte via les web services.
* Permettre de sécuriser les ressources, les données et les messages
* Permettre de chiffrer et déchiffrer les messages
* Permettre de protéger l’intégrité des données
  + 1. **Les cas d’utilisation**

1. Analyse
   1. Le moteur de recherche

La recherche simple est une fonctionnalité de base dans notre éditeur de texte, pour assurer la recherche nous allons nous baser sur le modèle de recherche d’information de base qui est version simplifié de celui utilisé par le moteur de recherche Google.

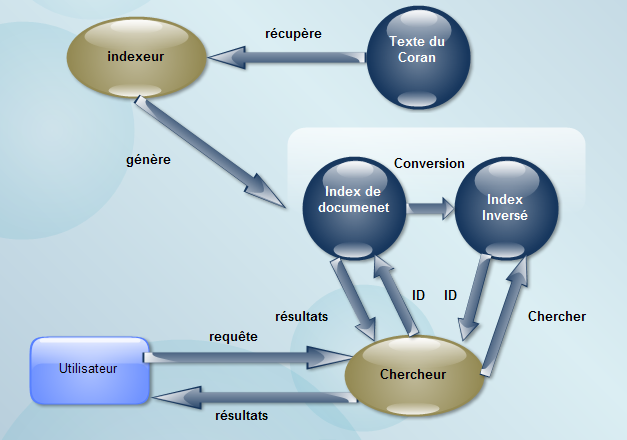


Figure 17:modèle de recherche de base

D’après la figure ci-dessus l’indexeur génère les index de base (index de document et index inversé), ses index seront utilisés pour optimiser la recherche, les figures ci-dessous illustrent des exemples sur les index de base, l’index de document qui sera converti en index inversé

|  |  |  |  |
| --- | --- | --- | --- |
| SOURA | AYA | AYA-ID | DOC-ID |
| الفاتحة | 1 | **بسم الله الرحمن الرحيم** | 1 |
| الفاتحة | 2 | **الحمد لله رب العالمين** | 2 |
| الفاتحة | 3 | **الرحمن الرحيم** | 3 |
| الفاتحة | 4 | **مالك يوم الدين** | 4 |
| الفاتحة | 5 | **إياك نعبد وإياك نستعين** | 5 |
| الفاتحة | 6 | **اهدنا الصراط المستقيم** | 6 |
| الفاتحة | 7 | **صراط الذين أنعمت عليهم غير المغضوب عليهم ولا الضالين** | 7 |
| البقرة | 1 | **الم** | 8 |
| البقرة | 2 | **ذلك الكتاب لا ريب فيه هدى للمتقين** | 9 |

Table 1:format d'un index de document

Nous allons illustrer la conception du moteur de recherche avec les différentes étapes

* + 1. L’exploration

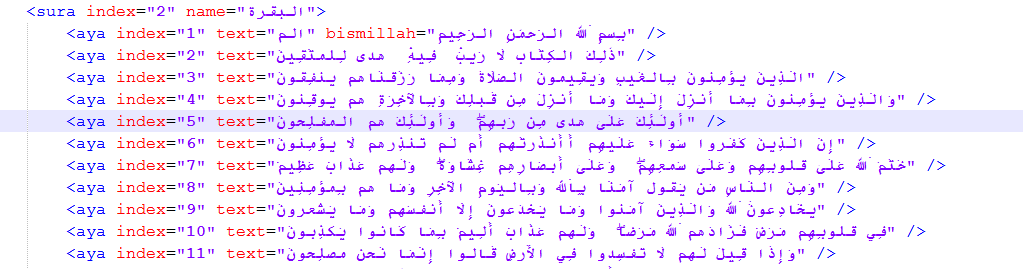
L’exploration est utilisée dans un environnement évolutif qui subit des changements comme le web, dans le cas du saint Coran, l’exploration est une phase inutile car les informations contenues dans le coran sont fixes.

* + 1. L’indexation

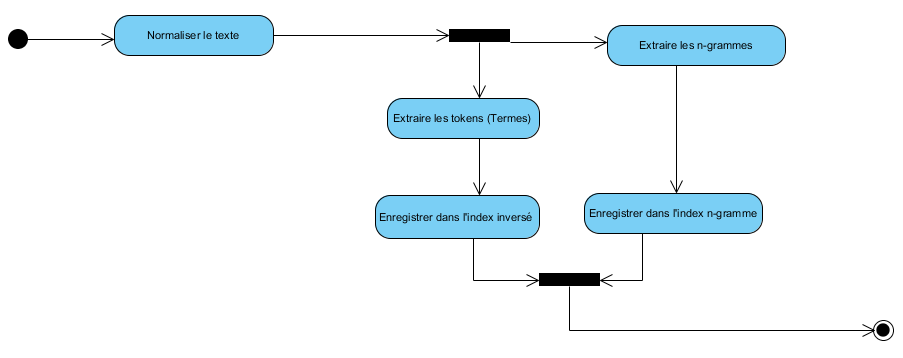
L’indexation est une étape très importante pour améliorer le temps de réponse et la qualité des résultats, car la force de la recherche réside dans les index utilisé pour accélérer la recherche et assister l’utilisateur, les principaux index que nous allons utiliser sont

* + - 1. Les étapes de l’indexation

Le point de départ de l’indexation est un document contenant le saint Coran décomposé en ayate avec un texte vocalisé et contenant tous les signes du Coran la figure ci-dessous montre les données à indexer (le texte original du Coran)



La recherche de texte dans le Coran avec la structure de la figure ci-dessous est de type séquentiel, pour cela des index doivent être utilisé, le diagramme ci-dessous illustre les différentes étapes d’indexation



**Normalisation** : Lorsque l’utilisateur introduit une requête de recherche, elle généralement non vocalisé, pour cela l’objectifs de la normalisation et d’enlever tous les signes spéciaux et ne garder que les lettres arabe pour avoir un langage commun entre la requête inséré et les index stocké

**بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ**

**Texte original**

**بسم الله الرحمن الرحيم**

**Texte normalisé**

**Tokenization :** cette opération consiste à décomposer le texte du Coran en unités appelées tokens qui sont généralement les mots qui constituent le texte, sans le cas du Coran il est difficile de séparer les affixes nominaux pour cela les tokens sont séparé seulement par des espaces.

**بسم الله الرحمن الرحيم**

**Texte normalisé**

**بسم**

**الله**

**الرحمن**

**الرحيم**

**Les Tokens**

**الل**

**لله**

**بسم**

**الر**

**Les n-grammes**

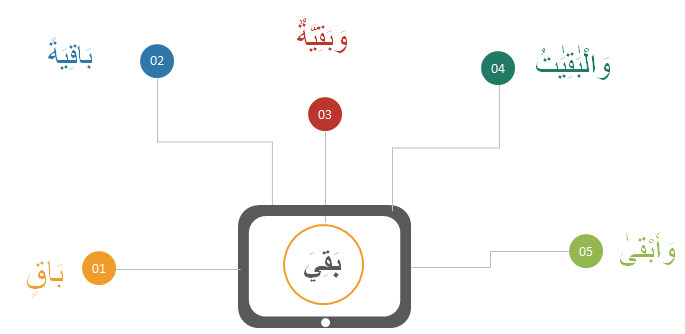
Les étapes ci -dessous seront appliqué sur tout le Coran à la fin obtenons l’index inversé et le n-gramme index

|  |  |
| --- | --- |
| DOC-id | n-gramme |
| 1.. | **بسم** |
| 1.. | **الله** |
| 1,3.. | **الرحمن** |
| 1,3.. | **الرحيم** |
| 1.. | **الحمد** |
| 1,2.. | **لله** |

|  |  |
| --- | --- |
| DOC-id | Terme |
| 1.. | **بسم** |
| 1.. | **الل** |
| 1,3.. | **لله** |
| 1,3.. | **الر** |
| 1.. | **لرح** |
| 1,2.. | **رحم** |

**Segmentation :** La segmentation dans la langue arabe est plus complexe que celle des autres langues, il y a plusieurs opérations à effectuer pour réduire les différentes formes et les regrouper dans un seul mot racine, la racine étant la plus petite unité qui rassemble les familles du mot dans la langue arabe.

La figure suivante illustre un exemple de la segmentation



D’après la figure ci-dessus pour la racine بَقِيَ il y a au moins cinq termes correspondant, pour assurer la segmentation, l’index n-gramme peut donner quelque terme qui ont la même racine ( بَاقِيَةً, وَبَقِيَّةٌ) pour assurer une bonne recherche par racine nous allons mettre créer un autre index (index de morphologie) qui associe à chaque terme ,la racine correspondante

|  |  |
| --- | --- |
| Racine | Terme |
| إسم | **بسم** |
| الله | **الله** |
| رحم | **الرحمن** |
| رحم | **الرحيم** |
| حمد | **الحمد** |
| الله | **لله** |

* + 1. L’interrogation

L’indexation se fait avant de mettre le moteur de recherche en marche contrairement à l’interrogation qui s’exécute après que l’utilisateur introduire la requête et termine par l’exposition des résultats à l’utilisateur.

**Etablissement de la requête**

Le processus de l’interrogation commence par l’insertion de la requête dans la barre de recherche, nous allons assister l’utilisateur dans sa recherche simple par l’auto-complétion de la requête en se basant sur les premiers caractères insérés dans la barre de recherche,

Cette option est assurée par l’index n-gramme.

**Traitement de la requête**

Le traitement de la requête diffère selon le type de recherche, la recherche simple, avancée et programmé, mais l’objectifs est le même il faut extraire un liste de mots clés normalisé et d’opérations qui influence sur les résultats obtenue, ces opération peuvent être :

Recherche simple, par champ, relations logique, intervalle, Boosting, synonyme, dérivations, fautes d’orthographe …etc

**La recherche**

Une fois la requête traité, le moteur charge les index selon le type de recherche, la recherche simple utilise l’index inversé, la recherche par de joker utilise l’index n-gramme, et l’index de morphologie pour chercher les dérivations …etc, chaque opération de recherche nécessite des index qui donnent de meilleurs résultats

|  |  |  |  |
| --- | --- | --- | --- |
| Opération de recherche | Index utilisé | Exemples de requête | Exemple de réponses |
| Simple | Inversé | **محمد** |  |
| Par terme | Inversé | **محمد, أحمد** |  |
| La suggestion | n-gramme, de morphologie | **حمد** |  |
| Les fautes d’orthographe | n-gramme | **كحمد** |  |
| Expression logique | Inversé | **أحمد or محمد** |  |
| Joker | Inversé  n-gramme | **\*حمد\*** |  |
| Dérivation | morphologie | **حمد** |  |

**Traitement des résultats:**

Une fois les versets correspondants à la requête du client sont sélectionnés, nous effectuons plusieurs opérations pour afficher les résultats :

**Classement des résultats :**

Le classement des résultats se fait par ordre de pertinence, en calculant le poids (weight) du résultat, la formule de calcul du poids dépend des opérations de recherche et l’occurrence d’apparition des termes recherchés.

Le classement peut ne pas représenter la pertinence des résultats par rapport à la requête, si nous nous basons sur l’ordre dans le Mushaf, l’ordre de révélation ou l’ordre alphabétique.

**Suggestion de mots clé alternatifs**

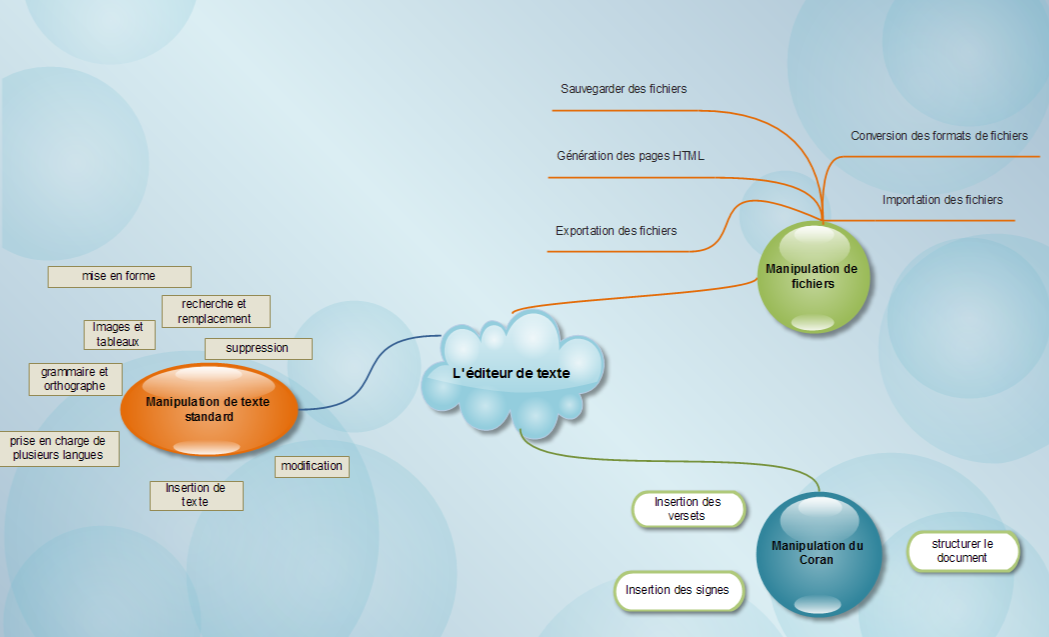
Cette option est exploitée lorsqu’il y a aucun résultat qui correspond à la requête de client en proposant des termes qui ressemblent celles de la requête, ou pour donner plus de termes lié à requête en se basant sur les index n-gramme et de morphologie.

**Extension vers d’autres documents**

Comme nous nous avons déjà cité, chaque verset est lié avec des informations supplémentaire, nous citons Les traductions, les récitations, le tafssīr.

* 1. L’éditeur de texte

L’éditeur de texte est le composant principal de notre application il est composé de plusieurs modules illustrés dans la figure ci-dessous



Comme nous le montrons dans la figure le composant éditeur de texte est subdivisé en lui-même en trois modules, qui sont :

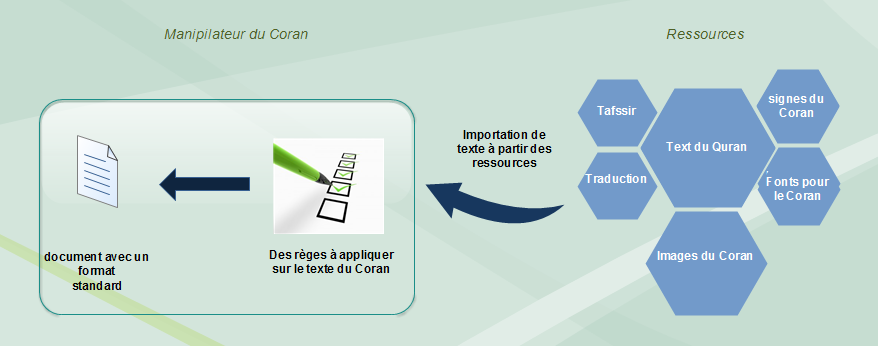
* Manipulateur de texte standard qui est un éditeur de texte ordinaire
* Manipulateur de fichiers permettant l’ouverture l’importation, et l’exportation des fichiers
* Manipulateur de texte Coranique qui permet l’insertion des différents composants lié aux Coran comme le texte avec les signes spéciaux, les images le tafssīr et la traduction

Nous allons nous focaliser dans la conception sur la manipulation de texte Coranique et la manipulation des fichiers.

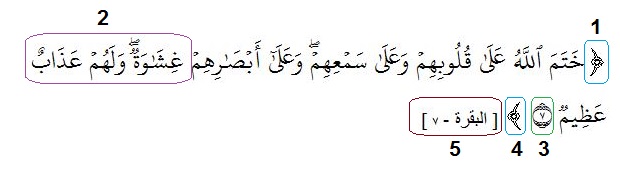
* + 1. Manipulation de texte coranique

Ce module couvre les limites de l’édition du texte Coranique, il permet d’insérer des versets Coranique authentiques sous format texte et qui supporte les signes marqué dans le script othmani:

* + - 1. Modèle de base



D’après le schéma ci-dessus le document Coranique est représenté sous un format standard qui peut être XML, HTML ou un autre standard basé sur XML, car nous n’allons pas nous contenté de la représentation des textes coraniques sur l’éditeur de texte mais aussi sur le transfert de différents texte via les web services.



La figure ci-dessus illustre les composants principaux trouvés dans une citation Coranique :

Le signe de début de verset

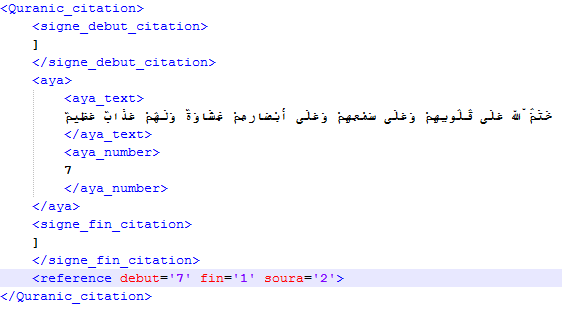
Une partie de verset, le verset commence de خَتَمَ et se termine par عَظِيمٞ .

Le numéro de verset, dans le Coran les chiffres sont représentés dans l’écriture arabe.

Le signe de fin de verset ou une suite de versets successifs.

Une référence de la citation contenant le nom de Soura et les numéros du premier et du dernier verset cités.

La majorité ses signes du Coran sont dans la deuxième partie (le text de verset) , L’exemple précédent ai illustré ci-dessous au format XML



Des règles doivent être appliquées sur les citations Coranique pour conserver les caractéristiques et le contenu du Coran.

* Il faut utiliser le clavier arabe pour écrire les citations Coranique.
* La direction du texte doit être de droite à gauche.
* Le style du script othmani doit être appliqué sur le texte du Coran
* Certains signes ne sont utilisés qu’avec le texte Coranique, exemple des signes de début et fin d’aya
* L’éditeur de texte doit empêcher la modification des versets, les opérations possibles sont l’insertion et la suppression des versets
* Ces règles sont appliquées aussi sur les textes liés au Coran comme les traductions et le tafssīr
  + - 1. Validation des citations de Coran

La validation consiste à vérifier l’intégrité des versets coraniques, elle peut être manuelle ou automatique avant d’enregistrer ou exporter le document, le hachage peut être une méthode efficace pour garantir l’intégrité des versets.

Schéma de validation

* + - 1. La suggestion
    1. Manipulation de fichiers

Ce module permet de manipuler les différents types de fichier (PDF, Word, texte et image) en permettant les opérations d’importation et exportation et sauvegarde de fichiers.

* + - 1. **Importation de texte**

Cette fonctionnalité permet aux utilisateurs d’importer le texte contenu dans plusieurs types de fichiers les plus utilisé, cette opération passe par plusieurs étapes :

* Détection du type de document : après le choix de fichier à importer nous commençons par la détection de types de fichier qui peut être un fichier PDF, Word, texte ou image
* Extraction de texte : dans cette étape en se basant sur la structure de document qui change d’un format à un autre, nous parcourons la totalité de fichier pour extraire le texte contenue dans le fichier
* Encodage Unicode : une fois on extrait le texte on le convertie en Unicode pour supporter le texte Coranique.
* Effectuer des transformations sur le texte : certain transformations doivent être effectué sur le texte pour qu’il soit compréhensible notamment la direction de texte
* Transformation en format http
  + - 1. Exportation de fichiers

Cette tâche est plus facile que la précédente car le texte à exporter est sous un format standard HTML, l’exportation se fait en deux étapes

* Analyser la page HTML et extraire le texte à partir de l’éditeur
* Détecter les différentes citations Coranique et déclencher le processus de validation des versets
* Générer au format de destination ou enregistrer dans l’espace de l’utilisateur.
  1. Les web services

Description des différents protocoles pour la mise en place des web services

L’édition de texte via les web services

La recherche via les web services

1. Conception

Diagramme de classe

# Partie 3 : Implémentation

1. Bibliographie

ABAR , A. La capitalisation du profil des experts du CENEAP sur la base de la classification des études. Ecole nationale superieure d'informatique, 2009.

ALKAHFLEJOURNAL.COM. Description du Coran. In*.*, 2014.

ALKHARASHI, I. A Web Search Engine for Indexing, Searching and

Publishing Arabic Bibliographic Databases. In*.*: King Abdulaziz City for Science and Technology 1999.

ALLAB, K. Classification Automatique de Documents XML. 2008.

ALTERNATIVETO. Alfanous. In*.*, 2014.

BABIN, G. AND M. LEBLANC. Les Web Services et leur impact sur le commerce B2B. 2003.

BALOUL, S. Développement d'un système automatique de synthèse de la parole à partir du texte arabe standard voyellé. Le Mans, 2003.

BOUDALI, F. Publication et découverte des web services pour le domaine du e-learning, mémoire de magister. INI Alger, 2007.

BOURNE, C. P. AND B. ANDERSON *DIALOG lab workbook: training exercises for the DIALOG® Information Retrieval Service*. Edtion ed.: DIALOG Information Retrieval Service, Lockheed Corp., 1981. ISBN 0933414072.

BOX, D., D. EHNEBUSKE, G. KAKIVAYA, A. LAYMAN, et al. Simple object access protocol (SOAP) 1.1. In*.*, 2000.

BRIN, S. AND L. PAGE The anatomy of a large-scale hypertextual Web search engine. Computer networks and ISDN systems, 1998, 30(1), 107-117.

BROWN, A. AND H. HAAS Web Services Glossary 2002.

CERAMI, E. *Web services essentials: distributed applications with XML-RPC, SOAP, UDDI & WSDL*. Edtion ed.: " O'Reilly Media, Inc.", 2002. ISBN 0596002246.

CHARTRON, G., S. DALBIN, M.-G. MONTEIL AND M. VÉRILLON Indexation manuelle et indexation automatique: dépasser les oppositions. Documentaliste, 1989, 26(4-5), 181-187.

CHELLI, A. Proposal of an Advanced Retrieval System for Noble Qur’an. Ecole nationale superieure d'informatique 2013.

CHELLI, A. AND M. DAHMANI. Développement d’un moteur de recherche et d’indexation

des documents Coraniques. Ecole nationale superieure d'informatique ESI, 2010.

DICODUNET. Sécurité informatique. In*.*, 2014.

FLEURISLAM.COM. Analyse historique du Coran In*.*, 2014.

GHERNAOUTI, S. *Sécurité informatique et réseaux - 4e édition: Cours avec plus de 100 exercices corrigés*. Edtion ed.: Dunod, 2013. ISBN 9782100701841.

GODART, D. *Sécurité informatique: risques, stratégies et solutions*. Edtion ed.: Edipro, 2002. ISBN 2930287217.

HADJ HENNI , M. Approche ontologique pour la modélisation sémantique, l’indexation et l’interrogation des documents coraniques. Ecole nationale superieure d'informatique, 2008.

ISLAMWEB.COM. L’histoire de la Révélation du Coran In*.*, 2014.

JENSEN, M., N. GRUSCHKA AND R. HERKENHÖNER A survey of attacks on web services. Computer Science-Research and Development, 2009, 24(4), 185-197.

KADIMA, H. AND V. MONFORT *Les Web services: techniques, démarches et outils XML, WSDL, SOAP, UDDI, Rosetta, UML*. Edtion ed.: Dunod, 2003. ISBN 2100065580.

KELLERT, P. AND F. TOUMANI Les web services sémantiques. CLR03], chapitre, 2003, 7, 93-106.

KINGFAHDBINABDULAZIZ.COM. Provision of copies of the Holy Quran. In*.*, 2014.

LACKEY, Z., H. DWIVEDI AND R. CANNINGS *Hacking sur le Web 2.0*. Edtion ed.: Pearson Education France. ISBN 2744040894.

LELU, A. AND C. FRANÇOIS. Information retrieval based on a neural unsupervised extraction of thematic fuzzy clusters. In *Les réseaux neuro-mimétiques et leurs applications. Journées internationales.* 1992, p. 93-104.

LIDDY, E. How a Search Engine Works. Searcher, 2001, 9(5).

MAHSSIN, M. S. *تاريخ القرآن الكريم*. Edtion ed.: الأصفهاني للطباعة بجدة دار, 1973.

MANNING, C. D., P. RAGHAVAN AND H. SCHUTZE *Introduction to Information Retrieval* Edtion ed.: Cambridge University Press, 2009. ISBN 0521865719

MORTEN, I. Indexing for Beginners. In*.*, 2013.

NEJJARI, Y. Recherche documentaire Sur Le Web. In*.*, 2007.

QURANCOMPLEX.GOV.SA. بشرى سارة: إطلاق خط الرسم العثماني – حفص رسمياً. In*.*, 2011.

RADHAMANI, G. AND G. R. K. RAO *Web services security and e-business*. Edtion ed.: IGI Global, 2007. ISBN 1599041707.

ROBERTSON, S. E., S. WALKER, M. BEAULIEU AND P. WILLETT Okapi at TREC-7: automatic ad hoc, filtering, VLC and interactive track. Nist Special Publication SP, 1999, 253-264.

SALTON, G. The SMART retrieval system—experiments in automatic document processing 1971.

SANAN, M. Étude des méthodes de la recherche d'information et de l'indexation sur les documents électroniques: cas de la langue arabe. 2008.

SAUVAGNAT, K. Modéle flexible pour la Recherche d’Information dans des corpus de documents semi-structurés. Université Paul Sabatier Toulouse, 2005.

SMAHI, Z. Sécurisation des Web Services SOAP contre les attaques par injection par la méthode Khi-2 (χ²). Ecole nationale supérieure d'informatique 2014.

SOURCEFORGE.JP. Arabeyes. In*.*, 2015.

SULLIVAN, B. XML Denial of Service Attacks and Defenses. In *MSDN.* 2009.

TANZIL. Quran Metadata. In*.*, 2014.

TANZIL.NET. In*.*, 2014.

W3C. Le W3C publie le Cryptage XML et la Transformation de Décryptage en recommandations. In*.*, 2002.

W3C. Security. In*.*, 2014.

W3SCHOOLS. SOAP Header Element. In*.*, 2014.

WEBAPPSEC. XML Entity Expansion In*.*, 2014.

YOUTIPS.COM. Police d’écriture. In*.*, 2014.

ZERROUKI, T. UN MODELE DE MUSHAF ELECTRONIQUE STANDARD. Ecole nationale superieure d'informatique, 2005.

الطيار, م. ب. س. 2008. المحرر في علوم القرآن. In *Proceedings of* 2008.

خضر, م. ز. التعامل مع القرآن الكريم في عصر المعلوماتية. In *مؤتمر كلية الآداب الخامس.* جامعة الزرقاء الأهلية, 2004, p. 23.