Boosting

It’s about improving the simple search query to do some specific search needs like search by phrase, intervals (numeric, alphabetic), logical relations (AND, OR, AND-NOT, AND-MAYBE …), regular expressions (jokers) and fielded search. Moreover, these can be enriched by adding new features or altering the old ones to fit the Arabic and Quran properties and to meet their specific needs.

Améliorations:

Rechercher sur les meilleures méthodes de calcul pour recueillir des statistiques coranique.

Quelles statistiques peut-on prélever ?

Avons-nous besoin des index manuels ?

Et comment le faire avec une meilleure précision ?

Script Uthmani ou Standard?

Feature Name: description, Objective, notes de customization + examples

Advanced Query

* Fielded Search: use fieldname in the query to search in a specific field. Helpful to search more information like surah names.
  + سورة: الفاتحة
    - Logical relations : to force the existence or absence of a keyword. The most known relations are: AND for conjunction, OR for disjunction and NOT for exception. The relations can be grouped using parenthesis.
  + (الصلاة - الزكاة ) + سورة:البقرة
* Phrase search: force the existence of all keywords in the same order .
  + ”الحمد لله“
* Interval search: used to search an interval of values in the numerical field. Helpful in fields like: Ayah ID, Page, Hizb, statistic fields.
  + رقم\_الآية:[1 إلى 5]
  + Regular expressions (Jokers): used to search a set of words that share some letters. This feature can be used to search a part of a word.
    - In latin , there is two Jokers used largely: ? : replace one letter

\*: replace an undefined number of letters

* + ب؟طة 🡺 بسطة, بصطة
  + \*نبي\* 🡺 نبي , النبيين ، الأنبياء ,,,

note:

These jokers are inefficient in Arabic because:

* the existence of vocalization symbols which they are not letters,
* Hamza(ء) letter have different forms (different Unicode emplacements)
* Boosting : used to boost the relevance factor of any keywords.
  + Ex : سميع^2 بصير
* Combining features : search using a combination of elementary features

”\*حمد لله“^2

Output improvments

* Pagination: dividing the results on pages
  + 10, 20,50… results per page
  + Sorting: sort the results by various criteria such as:
  + Score
  + Mushaf order
  + Revelation order
  + Numerical order of numeric fields
  + Alphabitical or Abjad order of alphabetic fields
  + A combination of the previous orders

note:

for the alphabetical order , must consider the real order of :

Hamza forms : ؤ ئ ء أ

Ta’ forms: ة ت

Alef forms : ى ا

* Highlight : used for distinction of searched keywords in Ayah

الحمد <style>لله </style>رب العالمين

Note: vocalization, the keyword searched can be different to the showed one

* Real time output: used to avoid the time of user waiting ,and show the results directly when retrieved

* Results grouping: the criteria can be used for grouping are:
  + By similar Ayahs [whats similar ayahs]
  + by Surahs
  + by subjects
  + by taffsir dependency
  + by revelation events
  + By Quranic examples
* Uthmani Script with Full diacritical marks:

**[ ۞ لَّقَدْ كَانَ فِى يُوسُفَ وَإِخْوَتِهِۦٓ ءَايَٰتٌ لِّلسَّآئِلِينَ]**

Suggestion systems

* Vocalized Spell Correction: offers alternatives for keywords when misspelled or appeared with a different form in Quran.
  + أبرَاهَام: إبْرَاهِيم

* Semantically Related Keywords: used as hints, generally based on domain ontologies.
  + يعقوب: يوسف، الأسباط، نبي...
  + Different vocalization: suggest all the possible vocalizations sorted by the frequencies
  + الملك : المَلِك ، المُلْك، المَلَك ...
* Collocated words: provide a completion based on the collocation statistics
  + سميع : سميع عليم، سميع بصير

الحمد: الحمد لله

* Different Significations: used to limit a keyword on only one of their meanings

رب: معنى 1 (إله)، معنى 2 (سيد)،

linguistic AsPECts

* Partial Vocalization Search: gives user the opportunity to specify some diacritics and not all, helpful to eliminate I3rab (إعراب).
  + مـَلـك to find مـَلـِك, مـَلـَك … and ignore مـُلـْك
* Multi-level Derivation : the Arabic words derivations can be devided to four levels :
  + - * Exact word: فأسقيناكموه
      * Word affixes removed (Lemma) : أسقينا
      * Stem: أسقى
      * Root : سقي
  + Example:
  + (word:سقي , level: root) to find **يَسْقُونَ, نَسْقِي, وَيَسْقِينِ , يُسْقَوْنَ , وَيُسْقَوْنَ, وَسُقْيَاهَا , وَأَسْقَيْنَاكُمْ , وَيُسْقَى , لَأَسْقَيْنَاهُمْ , وَنُسْقِيَهُ , فَأَسْقَيْنَاكُمُوهُ , فَسَقَى , السِّقَايَةَ, نُسْقِيكُمْ, فَيَسْقِي, سِقَايَةَ, تَسْقِي, سَقَيْتَ, يُسْقَى, اسْتَسْقَى, اسْتَسْقَاهُ, وَسُقُوا.**
  + (Word: أسقينا, level: lemma) to find **وَأَسْقَيْنَاكُمْ** , **لَأَسْقَيْنَاهُمْ, فَأَسْقَيْنَاكُمُوهُ**.
  + Specific-derivations: this is specification of the pervious feature .Since Arabic is fully flexional , it has lot of derivation operations .
  + Conjugation in Past tense of قال to find قالت, قال, قالوا, قلن ...
  + Conjugation in Imperative tense de قال to find قولوا , قل ...
* Word Properties Embedded Query: offers a smart way to handle words families by filtering using a set of properties like:

*root , lemma, type, part-of-speech,Verb mood, Verb form, gender, person, number, voice…etc.*

* + {جذر:**ملك** نوع:**اسم** عدد: **مفرد** }
* Numerical values substitution: this helps to retrieve numbers, even if appeared as words.
  + 309 replaces ثلاثمائة وتسعة
  + Consideration/Ignoring Spell faults : especially for the letters that usually misspelled like Hamza(ء); The Hamza letter is hard to write its shape since its writing is based on the its vocalization and the vocalization of its antecedent.

* + مءصدة replacesمؤصدة
  + ضحي replaces ضحى
  + جنه replaces جنة
* Uthmani writing way: offers the user the ability of writing words not as it appeared in Uthmani Mushaf .
  + بصطة replaces بسطة
  + نعمت replaces نعمة
* Pronoun reference resolving :
  + لا اله إلا هو, هو 🡺 الله

QUranic Options

* + Recitations marks retrieving: helpful for Tajweed scholars.
  + سجدة : نعم
  + سكتة:واجبة \_نوع
  + قلقلة:نعم
  + Structural options :

بنية داخلية منظمة فهو مقسّم إلى أجزاء والأجزاء إلى أحزاب والأحزاب إلى أنصاف و....إلى أن نصل إلى الآية وهناك تقسيمات أخرى كالتقسيم إلى صفحات...إلخ.

* + صفحة: 1
  + حزب : 60
* Translation Embedded Search
  + { text: mercy lang: english author: shekir }
* Similitudes search (المتشابهات)
  + متشابهة {55,13} == [ فَبِأَيِّ آلَاءِ رَبِّكُمَا تُكَذِّبَانِ] , 31 similitudes exactes
* Examples search (الأمثال)
  + مثل سورة:البقرة

**[ مَثَلُهُمْ كَمَثَلِ الَّذِي اسْتَوْقَدَ نَارًا فَلَمَّا أَضَاءَتْ مَا حَوْلَهُ ذَهَبَ اللَّـهُ بِنُورِهِمْ وَتَرَكَهُمْ فِي ظُلُمَاتٍ لَا يُبْصِرُونَ]**

SeMaNTIC QUeries

* + semantically related words

|  |  |  |
| --- | --- | --- |
| [Synonymy](http://en.wiktionary.org/wiki/Wiktionary:Semantic_relations#Synonymy) | Each listed synonym denotes the same as this entry. | ====Synonyms==== |
| [Antonymy](http://en.wiktionary.org/wiki/Wiktionary:Semantic_relations#Antonymy) | Each listed antonym denotes the opposite of this entry. | ====Antonyms==== |
| [Hypernymy](http://en.wiktionary.org/wiki/Wiktionary:Semantic_relations#Hypernymy) | Each listed hypernym is superordinate to this entry; This entry’s referent is a kind of that denoted by listed hypernym. | ====Hypernyms==== |
| [Hyponymy](http://en.wiktionary.org/wiki/Wiktionary:Semantic_relations#Hyponymy) | Each listed hyponym is subordinate to this entry; Each listed hyponym’s referent is a kind of that denoted by this entry. | ====Hyponyms==== |
| [Meronymy](http://en.wiktionary.org/wiki/Wiktionary:Semantic_relations#Meronymy) | Each listed meronym denotes part of this entry’s referent. | ====Meronyms==== |
| [Holonymy](http://en.wiktionary.org/wiki/Wiktionary:Semantic_relations#Holonymy) | Each listed holomym has this entry’s referent as a part of itself; this entry’s referent is part of that of each listed holonym. | ====Holonyms==== |
| [Troponymy](http://en.wiktionary.org/wiki/Wiktionary:Semantic_relations#Troponymy) | Each listed troponym denotes a particular way to do this entry’s referent. | ====Troponyms==== |
| [Coordinate term](http://en.wiktionary.org/wiki/Wiktionary:Semantic_relations#Coordinate_term) |  |  |

* + Syn(جنة ) pour la recherche de جنة, نعيم, فردوس …
  + Ant( جنة) pour la recherche de جحيم, سعير , جهنم, سقر …
  + Is( جنة) pour la recherche de عدن، فردوس
  + … (selon l’ontologie utilisée)
* Natural Questions: This means to offer the option of forming the search query as a form of an Arabic natural question . the Arabic question words are: هل؟من؟ ما؟ أين؟ متى؟ لم؟ كم؟ أيّ؟ لمن؟

~~من هم~~ الأنبياء؟

**[ إِنَّا أَوْحَيْنَا إِلَيْكَ كَمَا أَوْحَيْنَا إِلَى نُوحٍ وَالنَّبِيِّينَ مِنْ بَعْدِهِ وَأَوْحَيْنَا إِلَى إِبْرَاهِيمَ وَإِسْمَاعِيلَ وَإِسْحَاقَ وَيَعْقُوبَ وَالْأَسْبَاطِ وَعِيسَى وَأَيُّوبَ وَيُونُسَ وَهَارُونَ وَسُلَيْمَانَ وَآتَيْنَا دَاوُودَ زَبُورًا] -** النساء 163

~~ما هي~~ الحطمة؟

[ **نَارُ اللَّـهِ الْمُوقَدَةُ**] - الهمزة 6

أين غلبت/هزمت الروم؟

[**فِي أَدْنَى الْأَرْضِ وَهُمْ مِنْ بَعْدِ غَلَبِهِمْ سَيَغْلِبُونَ**] - الروم 3

كم مكث أصحاب الكهف؟

[ **وَلَبِثُوا فِي كَهْفِهِمْ ثَلَاثَ مِائَةٍ سِنِينَ وَازْدَادُوا تِسْعًا**] - الكهف 25

متى يوم القيامة؟

[**يَسْأَلُكَ النَّاسُ عَنِ السَّاعَةِ قُلْ إِنَّمَا عِلْمُهَا عِنْدَ اللَّـهِ وَمَا يُدْرِيكَ لَعَلَّ السَّاعَةَ تَكُونُ قَرِيبًا**] - الكهف 25

كيف يتشكّل الجنين؟

[**ثُمَّ خَلَقْنَا النُّطْفَةَ عَلَقَةً فَخَلَقْنَا الْعَلَقَةَ مُضْغَةً فَخَلَقْنَا الْمُضْغَةَ عِظَامًا فَكَسَوْنَا الْعِظَامَ لَحْمًا ثُمَّ أَنْشَأْنَاهُ خَلْقًا آخَرَ فَتَبَارَكَ اللَّـهُ أَحْسَنُ الْخَالِقِينَ**] -  المؤمنون 14

* Auto Vocalization: the absence of diacritics lead to the ambiguities that we’ve mentioned them in Problematic. This feature helps to pass over these ambiguities and try to resolve them before executing the search process.
  + رسول من الله 🡺 رَسُول مِنَ اللهِ
  + Proper nouns Search: lot of proper nouns are mentions clearly in Quran but some of them are just referred to implecitly . This feature is to search for proper nouns however mentioned : explicitly or implicitly.

بنيامين؟

[**إِذْ قَالُوا لَيُوسُفُ وَأَخُوهُ أَحَبُّ إِلَى أَبِينَا مِنَّا وَنَحْنُ عُصْبَةٌ إِنَّ أَبَانَا لَفِي ضَلَالٍ مُبِينٍ**] -  المؤمنون 14

أبو بكر/ الصديق؟

[**ثَانِيَ اثْنَيْنِ إِذْ هُمَا فِي الْغَارِ إِذْ يَقُولُ لِصَاحِبِهِ لَا تَحْزَنْ إِنَّ اللَّـهَ مَعَنَا**] -  التوبة 40

1. Statistical aims

* Vocalized word frequency:
  + - * Globally
      * In AYA,
      * In Sura,
      * In Hizb
* Root/Stem/Lemma frequency
* Listing collocated words

* + عدد\_الكلمات\_في الآية: 1
  + Le nombre d’apparitions des الدنيا et الآخرة avec tous ces dérivées

كم مرة وردت كلمة **"الله"**في سورة **"المجادلة"**؟

كم مرة وردت كلمة **" بحر"** في القرآن ؟ كلمة **" بحار"** تحسَب أيضا .

ما هي الكلمة الأكثر ورودا في القرآن ؟

ما هي الكلمات التي وردت أكثر من **1000** مرة  ؟

ما هي الكلمات التي وردت أكثر من **20** مرة في سورة **"البقرة"** ؟

ما هي الكلمات التي وردت أقل من **10** مرات؟

ما هي الكلمات التي وردت **20** مرة قبل سورة **"الكهف"**؟

ما هي أطول آية؟

+ Search for the new ones in your to-dos folder

## Alfanous API as a prototype

To validate our vision about search features on Quran, we work in implementing them through the API of Alfanous. Alfanous is an open source project that aims to build an indexing/search API in Quran and use it to develop different interfaces: Web, Desktop...etc. Alfanous API is pure python, prototyped from the library whoosh which is a fast, featureful full-text indexing and searching library implemented in pure Python. Programmers can use it to easily add search functionality to their applications and websites. Every part of how Whoosh works can be extended or replaced to meet specific needs exactly. The API helps the researchers and developers that working on Quran applications and want to take the benefits of search.

Alfanous API is:

* **Easy to use**, by virtue of a JSON output system that interacts with the API. This last can be used as a wrapping to oth
* **Easy to customize(prototyping),** because the simplicity and clarity of the code
* **Free Open Source,** developed under the license AGPL (Affero General Public License)

In this table, we’ll show some snippets of code for some available functions in Alfanous API

|  |  |
| --- | --- |
| Action | Code Sample |
| Import the Quranic Search Engine and Instantiate it | >>> from alfanous.main import QuranicSearchEngine  >>> QSE = QuranicSearchEngine(*"./indexes/main/"*) <\_\_main\_\_.QuranicSearchEngine instance at 0xb77d168c>  >>> QSE.OK  True |
| Make a search operation and browse the ayahs’ text | >>> results,\_ =QSE.search\_all(u*"الحمد"*)  >>> for r in results:  >>> print r[*"aya"*]  الحمد لله رب العالمين |
| Show more information for each result: surah *name, ayah* number, and ayah *vocalized text.* | >>> for r in results:  >>> print r[*"sura\_name"*],*":"*,r[*"aya\_id"*],*":"*,r[*"aya\_"*]  الفاتحة:2:الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ |
| Show the terms generated from a query | >>> \_,terms =QSE.search\_all(u*"الحمد"*)  >>> print terms  [(*"aya"*,u*"الحمد"*)] |
| use the terms to highlight the ayah text | >>> Qhighlight(u*"الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ"*, [u*"الحمد"*, u*"لله"*])  <span style=*"color:red;font-size:100.0%"*><b>الْحَمْدُ</b></span> <span style=*"color:green;font-size:100.0%"*><b>لِلَّـهِ</b></span> رَبِّ الْعَالَمِينَ |
| Show the estimated search time and the number of results | >>> print *"time=%f,number=%d"* % (results.runtime,len(results))  time=0.100318,number=160 |
| Sort the results in a different order, by example: Mushaf order | >>> results,\_ =QSE.search\_all(u*"الله"*,sortby=*"mushaf"*) |
| Suggest in the case of missing terms | >>> for key, value in QSE.suggest\_all(u*" عاصمو "*).items():  … print key, *":"*, *","*.join(value)  عاصمو : عاصم |

In the next table, we’ll mention some specific queries already implemented in the API:

|  |  |
| --- | --- |
| Feature | Syntax |
| simple search | سماكم |
| derivation level 1 | >سماكم |
| derivation level 2 | >>سماكم |
| joker ? | ال؟لك |
| joker \* | \*نبي\* |
| phrase | *"رب العالمين"* |
| spell errors | %المأصدة |
| relation logic And | الصلاة + الزكاة |
| relation logic Or | الصلاة | الزكاة |
| relation logic AndNot | الصلاة - الزكاة |
| Synonyms | ~السعير |
| fielded search | sura:الفاتحة |
| interval | sura\_id:[112 to 114] |
| vocalization | *'المَلك'* |
| word properties tuple | {ملك,فعل} |
| buckwalter | *fawoqa* |

# Related Works

There is lot of web applications and softwares that offer the service of search in Quran but most of them are simple search using exact words. We’ll mention some applications and websites that implemented some special search features.

* Alawfa search engine is a website that offers search in Quran and Hadith. He has a fast retrieving method. It offers the highlight, pagination. The main disadvantage of Alawfa is that it doesn’t respect the specifics of Arabic and Quran. It searches only as exact word or as a part of word.
* Al-Monaqeb alqurany (المنقّب القرآني) is a search service in Quran included in the website [**www.holyquran.net**](http://www.holyquran.net)**.** It offers a board for advanced search options that contains many useful options :
  + Transliteration
  + Ignore/Consider Hamza spelling errors
  + Limit the search in an interval of ayahs and surahs
  + Search using Arabic words root
  + Highlight & Pagination

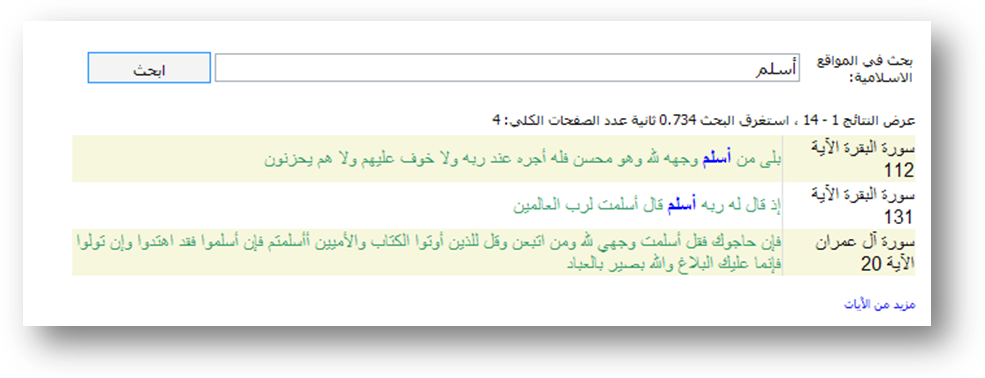
Al-Monaqeb offers no suggestions in the case of no results.

* Zekr application is a well known Quranic application aimed for browsing and studying the ayahs of Quran. It offers the browsing of Quran based on the different structures: Hizb, Sura, and Ayah. It offers also some advanced search features like search using Arabic roots , sorting in different orders: relevance.

Beta.quran.com

* 1. الأوفى

محرّك بحث مختص بالقرآن الكريم والحديث الشريف والمواقع الإسلامية ويعتبر نفسه الأول في هذا المجال...ما يميّزه عن باقي المواقع الأخرى التي توفر البحث في القرآن الكريم أنه أنشئ لهذا الغرض بالذات وبني على أسس محركات البحث.



يحمل الأوفى بعضا من الميزات الجيدة مثل استكشاف المواقع واقتراح الكلمات والتمييز..لكنها تبقى مزايا بسيطة بالمقارنة بما تملك محركات البحث الأجنبية. لا يوفر الأوفى لوحة بحث متقدّم بل يعتمد على التلقائية وهذا جيد لكنه لا يساعد الباحثين الذين يفضلون استعمال خيارات محدّدة دون أخرى.التحليل الصرفي للكلمات المدخلة سيء وشبه معدوم إضافةً إلى التعامل السيئ مع الحركات .اهتم الأوفى بعرض نتائج البحث في المواقع الإسلامية وتساهل في التعامل مع نتائج البحث في القرآن الكريم كأنها نتائج ثانوية.

موقع الأوفى يعمل الآن بإصدار تجريبي وفي انتظار أول إصدار كامل نرجو أن يتدارك نقاط الضعف الموجودة في الإصدار الحالي. (الأوفى، 2010)

* 1. المنقّب القرآني

هو عبارة عن خدمة للبحث في القرآن مدرجة في موقع القرآن الكريم ([www.holyquran.net](http://www.holyquran.net/)).يتميّز بلوحة بحث متقدّم بها عدة خيارات كتجاهل شكل حامل الهمزة والبحث بتهدئة الحروف الإنجليزية والبحث في مجال من الآيات أو السور...ولا يوفّر التحليل الصرفي إطلاقا بل يكتفي بالبحث بالجذر وتخيير المستخدم بين البحث ككلمة مستقلة أو كجزء من كلمة أخرى.

عرض النتائج في المنقّب جيد بعض الشيء حيث يعرض زر للتفسير مع كل آية مع تمييز الكلمات المنقّب عنها وتوزيع النتائج على صفحات.لا يعتمد المنقّب على التّلقائية ولا يقدّم أي اقتراحات فيبقى المستخدم كالأعمى في حالة كتابة الكلمة بوجه مغاير للوجه المحفوظ بها (مثلا: الكتابة العثمانية أو أخطاء الكتابة) .

ما يعاب عليه هذا الموقع أيضا هو اعتماده على مراجع شيعية للتفسير. (المنقب\_القرآني، 2010)

