مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (37) العدد (6) 2015
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (37) No. (6) 2015

# بناء نواة محرك بحث قادر على التحكم بالبحث وفهرسة المواقع

**سرار حمود***

**(تاريخ الإيداع 28 / 4 / 2015. قُبِل للنشر في 5/ 11 / 2015)**

## □ ملخّص □

يتضمن هذا البحث بناء نواة محرك بحث يمكنه العمل ضمن شبكة الانترنت ، قادر على التحكم بالبحث عن معلومات بمجالات محددة وفهرسة مواقع معينة .

تم في هذا البحث دراسة مسألة البحث عن المعلومات عبر الانترنيت و نظم استرجاع المعلومات وأنواع محركات البحث و المعماريات الأساسية لبناء المحركات و من ثم اقتراح معمارية محرك بحث يصلح نواة لمحرك البحث المرغوب وتحديد المخطط النهائي لمعمارية محرك البحث حيث تم بناء مقاطع محرك البحث و إجراء الاختبارات والنتائج.

**الكلمات المفتاحية :** محرك البحث – العنكبوت – المفهرس – نظام الاسترجاع .

* قائمة بالأعمال– قسم هندسة الحاسبات و التحكم الآلي– كلية الهندسة الميكانيكية و الكهربائية– جامعة تشرين– اللاذقية– سورية.

مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (37) العدد (6) 2015
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (37) No. (6) 2015

# Building a kernel of Search Engine overrule in search and indexing

**Serar Hammoud***

**(Received 28 / 4 / 2015. Accepted 5 / 11 / 2015)**

## □ ABSTRACT □

This research designs web search engine kernel overrule in searching of specific fields and indexing indicated sites.

This research contain information about search in web , retrieval system , types of search engines and basic architectures of building search engines .It suggests search engine architecture kernel of dedicated search engine to do final planner of search engine architecture ,and build parts of search engine and execute test to get results .

**Keywords** : Search Engine, Crawler, Indexer, Retrieval System

* **Academic Assistant, Department Computer and Automatic Control Engineering, Faculty of Mechanical and Electrical Engineering, Tishreen University, Lattakia, Syria.**

## مقدمة:

تنمو شبكة الانترنت بسرعة كبيرة، وتشكل خزاناً معلوماتياً هائلاً يضاف إليها ملايين الصفحات يومياً، لذلك لابد من إيجاد الحلول لمسألتين: مسألة زيادة حجم المعلومات من جهة، ومسألة تنظيم الوصول والبحث عن هذه المعلومات من جهة ثانية، من هنا ظهرت الحاجة إلى تصميم وسائل للوصول إلى هذه المعلومات بسرعة وسهولة بهدف توفير الوقت، من هذه الوسائل استخدام محركات البحث، ومحرك البحث هو كيان منطقي على شكل برنامج مصمم للمساعدة في تسهيل البحث عن المعلومات المطلوبة المتاحة على شبكة الإنترنت بسهولة ويسر حيث يعيد قائمةً بالمراجع التي توافق تلك المعلومات خلال فترة زمنية محددة. [1]،[2]

يمكن بناء محركات البحث بطرق عديدة، أشهرها:

- محركات بحث (عامة/خاصة).
- محركات بحث متعددة Meta search engine

تعتمد الفئة الاولى من محركات البحث على محرك بحث وحيد يمكن توجيه استعماله لينجز عملية البحث عن جميع المواضيع، أو تقنين عملية البحث ضمن مجال مخصص من المواضيع.

أما محركات البحث المتعددة، فهي تعتمد على مجموعة من محركات البحث، تشكل نظام بحث مركب، ويمكن أن تعتمد على عدة محركات بحث عامة يصل عددها إلى أكثر من عشرين محركا في وقت واحد. [3]،[4]

كذلك يمكن بناء محركات البحث لتعمل وفق قاعدة الكيانات والعلاقات ونميز فئتين هما:

- محركات بحث نصية، تطابق الطلب، كحروف لغوية، دون الغوص في معناها.
- محركات بحث دلالية semantic search engine، تهتم بالمعنى اللغوي للكلمات أو العبارات، و وصولا إلى هذا الهدف.

## أهمية البحث و أهدافه :

إن هدف البحث هو بناء نواة محرك بحث نصي متخصص، تم اقتراح معماريته، يتعامل مع قاعدة بيانات أصغر بكثير مقارنة بمحركات البحث الدلالية وبالتالي يحتاج زمنا أقصر للبحث واسترجاع المعلومات، ويتميز بتخصصه وقدرته على التحكم بالبحث عن معلومات بمجالات محددة، مع إمكانية حجب البحث عن مواضيع معينه، وإضافة المواقع وفهرستها، مع إمكانية التحكم بفهرسة روابط محددة ومنع فهرسة الروابط غير المرغوبة، واسترجاع أحدث النسخ عن المعلومات المطلوبة، بالإضافة إلى الحصول على إحصائيات لمعرفة وقت وعدد مرات تكرار البحث.

و من بين الدراسات السابقة الدراسة التالية،

Lipetz,Ben-Ami,Information retrieval systems, american library association,2013

التي تحدثت عن أنظمة استرجاع المعلومات بشكل عام، بينما تميزت دراستنا بتخصص البحث بمجال معين.

و الدراسة التالية،

Christopher D. Manning, Prabhakar Raghavan,Hinrich Schütze," Introduction to Information Retrieval", Springer, German,2009،

والتي تحدثت عن أنظمة الفهرسة واسترجاع المعلومات بشكل عام، بينما تم تزويد محرك البحث المقترح بقائمة حجب للكلمات و الروابط غير المرغوبة.

**طرائق البحث و مواده :**

لقد تم الاعتماد في بناء محرك البحث على مجموعة متنوعة من أدوات التطوير البرمجية، ومن أهم الأدوات المستخدمة في تصميمه، هو نظام إدارة قواعد البيانات MySql التي تتمتع بالعديد من الميزات، وكذلك لغة PHP المتعددة الخصائص والمستعملة لتطوير برامج الويب المختلفة.

إن بناء محرك البحث يتضمن بناء كل من العنكبوت، وقسم الفهرسة، والاسترجاع. [5]،[6]،[7]

يتم بناء العنكبوت بهدف إيجاد صفحات جديدة على الويب وإضافتها، أما الفهرسة فإن هدفها أن تأخذ المعلومات من العنكبوت وتعمل على تصنيفها ضمن فهارس مرتبة ومقسمة بترتيب معين لتسهيل تنفيذ عملية البحث عليها، أما بناء الاسترجاع فيقوم على تلقي الاستعلام والبحث ضمن الفهرس للمطابقة وتقديم النتائج الموافقة للاستعلام. [8]، [9].

**منهجية بناء العنكبوت المتبعة:**

انطلاقاً من تحليل عمل العنكبوت المبين في الشكل (1) يمكن تقسيم عملية البناء، إلى الخطوات التالية :

1- تخزين رابط صفحة البداية من أجل الفهرسة.

2- قراءة محتويات هذه الصفحة للبحث عن روابط موجودة في هذه الصفحة.

3- عند وجود رابط جديد إضافة هذا الرابط للفهرسة.

4- العودة إلى النقطة الثانية.

**منهجية بناء المفهرس المتبعة:**

إن تنفيذ المفهرس المبين في الشكل (1) يتم بالخطوات التالية :

1- قراءة محتويات الصفحة ووضعها في قاعدة البيانات مع الرابط الذي تنتمي إليه.

2- حساب وزن كل كلمة مع تجاهل بعض الكلمات الشائعة بكثرة والتي تملك أهمية قليلة جداً في تحديد المستندات المطابقة للاستعلام وتخزينها في قاعدة البيانات ويتم تعيين الوزن بحيث يكون مساوياً لعدد مرات ورود الكلمة في الصفحة.

3- تنتهي مهمة الفهرسة للصفحة الحالية ويتم فهرسة الصفحة التالية التي تم الحصول عليها من قائمة الروابط في بناء العنكبوت.

**منهجية بناء الاسترجاع المتبعة:**

إن تنفيذ الاسترجاع المبين في الشكل (2) في محرك البحث يتم بالخطوات التالية :

1 – إدخال اسم الموضوع المطلوب البحث عنه.

2- يتم فلترة كلمة البحث والتخلص من مواضيع البحث غير المرغوبة.

3- البحث في جداول قاعدة البيانات عن الكلمات التي تطابق اسم موضوع البحث وترتيب النتائج.

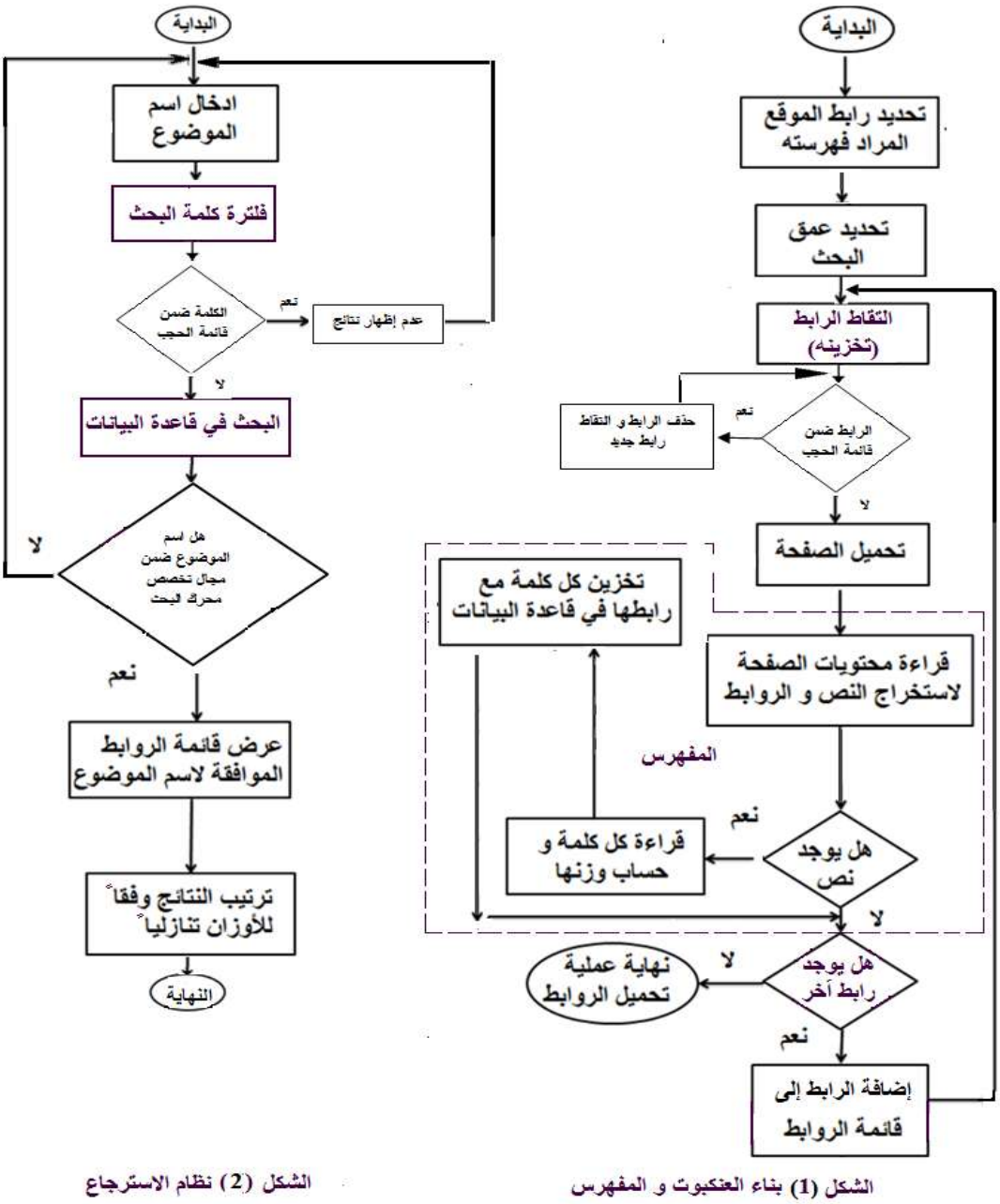
**معمارية محرك البحث المتخصص المقترح:**

يتعامل محرك البحث المصمم مع الطلب وفق قواعد البحث المقترحة، حيث يتم البحث عن الكلمة أو النص بشكل حرفي، كحروف لغوية ويظهر النتائج التي تتضمن تلك الحروف اللغوية.

لقد تم بناء محرك بحث مكون من مجموعة من المركبات المرتبطة بعضها ببعض، بنظام منطقي موضح عبر المخطط المنهجي المبسط المبين في الشكل (3).

يغطي المخطط المراحل الأساسية لبناء محرك البحث، حيث يقلع في البداية لتحديد عمق البحث ثم التقاط الرابط الذي نريد فهرسته وتحميل صفحة ذلك الرابط، بعد ذلك يتم قراءة محتويات الصفحة لاستخراج النص و الروابط، ثم يتم التأكد من وجود النص وفي حالة الإيجاب تتم قراءته كلمة كلمة مع حساب أوزان الكلمات وتخزينها مع الرابط الموافق في قاعدة البيانات وبهذا تكون قد تمت عملية الفهرسة. أما في حالة النفي لوجود النص يتم التأكد من وجود روابط أخرى، وفي حالة النفي تنتهي عملية تحميل الروابط، أما في حالة الإيجاب يتم وضع رابط الصفحة (الجديد) في قائمة الروابط وتحميل صفحة الرابط التالي (الجديد).

عند البحث عن موضوع معين ضمن مجال تخصص محرك البحث يتم البحث بقاعدة البيانات عن الروابط التي توافق اسم الموضوع وترتيبها تنازلياً وإيجاد الزمن المستغرق بالبحث، وفي حال عدم وجود روابط توافق اسم الموضوع أو إذا كان الموضوع خارج مجال تخصص محرك البحث فلا تظهر نتائج، حيث إن محرك البحث المتخصص المقترح يتطلب تصميم قائمة محددة لحجب الكلمات غير المرغوب إيجادها في مرحلة الاسترجاع، وحجب الروابط غير المرغوب فهرستها في مرحلة العنكبوت.




الشكل (1) بناء العنكبوت و المفهرس الشكل (2) نظام الاسترجاع

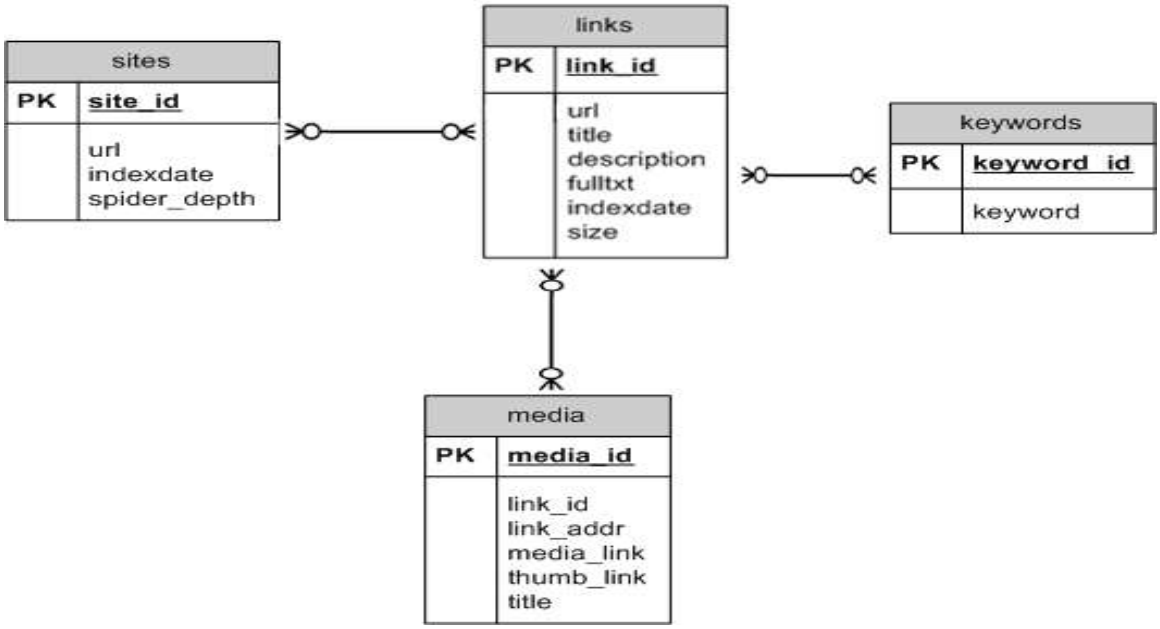
**الشكل (3) المخطط المقترح لمعمارية برنامج محرك البحث**

**تحليل النظام :**

تم تحليل النظام لمحرك البحث باستخدام لغة UML والتي تقدم لنا صيغة نمذجة رسومية لوصف العناصر، وقد تم تحليل النظام اعتماداً على الخطوات المبينة في الجداول( 1) و(2) و(3) و(4) و(5) و(6) و(7) و(8) و(9) و(10) ، والأشكال (4) و(5).

**الجدول (1) جدول كيانات النظام**

| اسم الكيان | اسم جدول الكيان | خصائص الجدول |
|---|---|---|
| كيان المواقع | sites | site_id: تمثل رقم الموقع الذي تم فهرسته.<br>url: رابط الموقع الذي تم فهرسته.<br>indexdate: تاريخ فهرسة الموقع.<br>spider_depth: عمق البحث. |
| كيان الكلمات | keywords | keyword_id: تمثل رقم الكلمة التي تم قراءتها و تخزينها بقاعدة البيانات.<br>keyword: الكلمة التي تم قراءتها و تخزينها بقاعدة البيانات. |
| كيان الروابط | links | link_id: رقم الرابط الذي تم فهرسته.<br>site_id: رقم الموقع الذي يتبع له الرابط .<br>url: الموقع الذي تمت فهرسته.<br>title: عنوان الرابط الذي تمت فهرسته.<br>description: الوصف المختصر للرابط الذي تمت فهرسته.<br>fulltxt: النص الكلي للرابط الذي تمت فهرسته.<br>indexdate: تاريخ فهرسة الرابط.<br>size: حجم محتوى الرابط الذي تمت فهرسته. |
| كيان كلمة الرابط | link_keyword | link_id: ويمثل رقم الرابط الذي تم فهرسته.<br>keyword: ويمثل الكلمة التي تم قراءتها و تخزينها بقاعدة البيانات.<br>weight: وزن الكلمة حسب عدد مرات تكرارها في محتوى الرابط.<br>hits: عدد مرات تكرار وجود الكلمة في محتوى الرابط.<br>indexdate: تاريخ فهرسة الرابط. |
| كيان الصور | media | media_id: تمثل رقم الصورة.<br>link_id: رقم الرابط الذي يتضمن الصورة.<br>link_addr: عنوان الرابط الذي يتضمن الصورة.<br>media_link: رابط الصورة.<br>thumb_link: موقع الصورة المصغرة التي تم تخزينها بقاعدة البيانات.<br>title: عنوان الصورة التي تم تخزينها بقاعدة البيانات. |



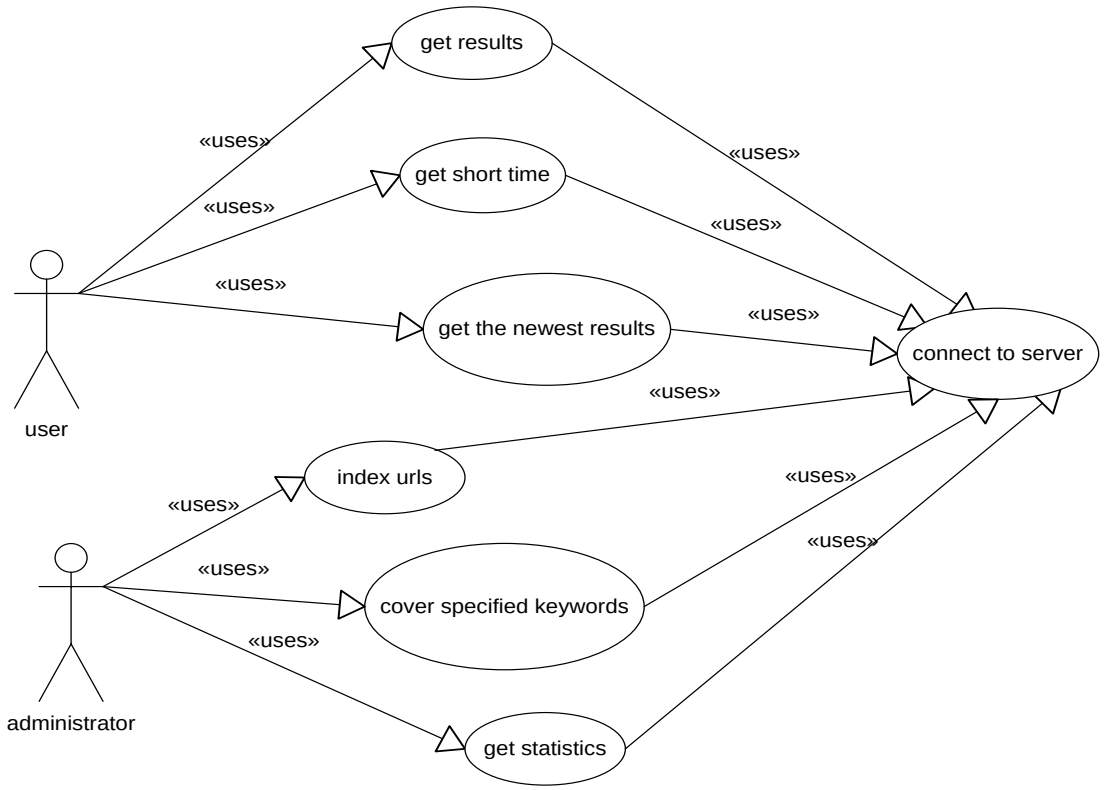

**الشكل (4) مخطط ERD**

**الجدول (2) متطلبـات النظام**

| |
|---|
| تخصيص البحث عن معلومات بمجالات محددة . get results |
| اختصار زمن البحث عن المعلومات (حسب الإمكانات المتاحة). get short time |
| استرجاع أحدث النسخ عن المعلومات المطلوبة.get the newest results |
| إضافة المواقع وفهرستها.index urls |
| حجب الكلمات غير المرغوبة أثناء البحث أو الفهرسة.cover specified keywords |
| إنشاء إحصائيات لمعرفة وقت وعدد مرات تكرار البحث عن كلمات البحث و الروابط. get statistics |

**الجدول (3) مستخدمي النظام**

| اسم المستخدم | الوصف |
|---|---|
| الباحث | يقوم بالبحث عن أحدث المعلومات المتخصصة بمجالات محددة بزمن قصير |
| المدير | إضافة المواقع وفهرستها وإمكانية حجب البحث عن مواضيع غير مرغوبة ومنع فهرسة رابط الموقع الذي يحوي كلمات غير مرغوبة وفهرسة الرابط الذي يحوي على كلمات مرغوبة ومعرفة زمن وعدد مرات تكرار البحث عن المواضيع والروابط. |




**الشكل (5) مخطط حالات الاستخدام (Use case diagram)**

**الجدول (4) حالة الاستخدام الحصول على النتائج**

| اسم حالة الاستخدام | وصف موجز | الفاعل | الشروط السابقة |
|---|---|---|---|
| الحصول على النتائج | تسمح هذه الخدمة للباحث تخصيص البحث عن معلومات بمجالات محددة | الباحث | لا توجد شروط مسبقة |
| التدفق الرئيس | | | التدفقات البديلة |
| المستخدم | النظام | | يقوم الباحث بإدخال كلمة بحث غير موجودة بقاعدة البيانات فيعرض النظام رسالة بعدم وجودها |
| في حالة البحث عن نص أو صورة يقوم الباحث بإدخال كلمات البحث المطلوبة | | | |
| | يقوم النظام بعرض النتائج التي حصل عليها من المخدم بأجزاء من الثانية و يظهر الزمن الذي استغرقته عملية البحث | | |

**الجدول (5) حالة الاستخدام اختصار زمن البحث**

| اسم حالة الاستخدام | وصف موجز | الفاعل | الشروط السابقة |
|---|---|---|---|
| اختصار زمن البحث | تسمح هذه الخدمة للباحث بالحصول على زمن بحث قصير (حسب الإمكانات المتاحة) | الباحث | لا توجد شروط مسبقة |
| التدفق الرئيس | | | التدفقات البديلة |
| المستخدم | النظام | | يقوم الباحث بإدخال كلمة بحث غير موجودة بقاعدة البيانات فيعرض النظام رسالة بعدم وجودها |
| في حالة البحث عن نص أو صورة يقوم الباحث بإدخال كلمات البحث المطلوبة | | | |
| | يقوم النظام بعرض النتائج التي حصل عليها من المخدم بأجزاء من الثانية و يظهر الزمن الذي استغرقته عملية البحث | | |

**الجدول (6) حالة الاستخدام استرجاع أحدث النسخ**

| اسم حالة الاستخدام | وصف موجز | الفاعل | الشروط السابقة |
|---|---|---|---|
| استرجاع أحدث النسخ | تسمح هذه الخدمة للباحث باسترجاع أحدث النسخ عن المعلومات المطلوبة. | الباحث | لا توجد شروط مسبقة |
| التدفق الرئيس | | | التدفقات البديلة |
| المستخدم | النظام | | يقوم الباحث بإدخال كلمة بحث غير موجودة بقاعدة البيانات فيعرض النظام رسالة بعدم وجودها |
| في حالة البحث عن نص أو صورة يقوم الباحث بإدخال كلمات البحث المطلوبة | | | |
| | يقوم النظام بعرض أحدث النتائج عن المعلومات المطلوبة التي حصل عليها من المخدم | | |

**الجدول (7) حالة الاستخدام فهرسة المواقع**

| اسم حالة الاستخدام | وصف موجز | الفاعل | الشروط السابقة |
|---|---|---|---|
| فهرسة المواقع | تسمح هذه الخدمة للمدير بإمكانية إضافة المواقع وفهرستها مع إمكانية حجب الكلمات غير المرغوبة | المدير | لا توجد شروط مسبقة |
| التدفق الرئيس | | | التدفقات البديلة |

| المستخدم | النظام | يقوم المدير بإدخال موقع غير موجود أو خاطئ فيتم عرض رسالة خطأ |
|---|---|---|
| يقوم المدير بالدخول إلى إعدادات البرنامج ثم إلى تبويب الفهرسة و من ثم يقوم بإدخال رابط الموقع المراد فهرسته و يحدد عمق الفهرسة و يختار فيما إذا أراد أن تتم فهرسة الموقع بفهرسة جميع روابط الموقع التابعة لنفس مجال الموقع أو فهرسة جميع روابط الموقع بغض النظر عن المجال | | |
| | يقوم النظام بعرض نتائج الفهرسة التي حصل عليها من المخدم | |

**الجدول (8) حالة الاستخدام حجب الكلمات غير المرغوبة**

| اسم حالة الاستخدام | وصف موجز | الفاعل | الشروط السابقة |
|---|---|---|---|
| حجب الكلمات غير المرغوبة | إمكانية حجب البحث عن مواضيع غير المرغوبة و منع فهرسة رابط الموقع الذي يحوي كلمات غير مرغوبة و فهرسة الرابط الذي يحوي على كلمات مرادة. | المدير | لا توجد شروط مسبقة |
| التدفق الرئيس | | | التدفقات البديلة |
| المستخدم | النظام | | يقوم المدير بإدخال موقع غير موجود أو خاطئ فيتم عرض رسالة خطأ |
| يختار المدير فيما إذا أراد حجب فهرسة المجالات غير المرغوبة و يختار أيضا فيما إذا أراد حجب فهرسة رابط الموقع الذي يحوي كلمات غير مرغوبة أو فهرسة الرابط الذي يحوي على كلمات مرغوبة و إذا أراد حجب البحث عن كلمات غير مرغوبة | | | |
| | يقوم النظام بعرض نتائج الفهرسة التي حصل عليها من المخدم | | |

**الجدول (9) حالة الاستخدام الحصول على إحصائيات**

| اسم حالة الاستخدام | وصف موجز | الفاعل | الشروط السابقة |
|---|---|---|---|
| الحصول على إحصائيات | تسمح هذه الخدمة للمدير بالحصول على زمن وعدد مرات تكرار البحث عن المواضيع و الصور و الروابط | المدير | لا توجد شروط مسبقة |
| التدفق الرئيس | | | التدفقات البديلة |
| المستخدم | النظام | | لا يوجد تدفقات |

| المدير | | بديلة |
|---|---|---|
| | يقوم النظام بعرض نتائج الإحصائيات المرادة | |

| الجدول (10) حالة الاستخدام الاتصال مع المخدم | | | |
|---|---|---|---|
| اسم حالة الاستخدام | وصف موجز | الفاعل | الشروط السابقة |
| الاتصال مع المخدم | تسمح هذه الخدمة للبرنامج بالاتصال مع المخدم للحصول على المعلومات | جميع حالات الاستخدام التالية الحصول على النتائج ، اختصار زمن الباحث ، استرجاع أحدث النسخ ، فهرسة المواقع | أن يتم طلب الاتصال من إحدى الخدمات السابقة |
| التدفق الرئيس | | | التدفقات البديلة |
| المستخدم | | النظام | في حال فشل الاتصال يعرض النظام رسالة خطأ توضح المشكلة |
| المدير | | يقوم النظام بفتح اتصال بالانترنت | |
| | | يرسل النظام المعلومات وينتظر النتائج | |
| | | يعيد النتائج التي تم طلبها | |

**بناء كود برنامج محرك البحث**

تم استعمال لغة php وقاعدة البيانات mysql في بناء مختلف المقاطع البرمجية للمحرك ، نبين لاحقا بعض المقاطع البرمجية المستخدمة في بناء محرك البحث:

1- المقطع البرمجي المستخدم من أجل الاتصال مع قاعدة البيانات :

```
$db_con     = db1_connect() ;
$success    = mysql_select_db ($database1, $db_con);
```

2 - المقطع البرمجي المستخدم من أجل الحصول على النتائج النصية :

```
 if ($media_only==0 ) {
 $text_results = get_text_results($query, $start, $results);
$ip = $_SERVER['REMOTE_ADDR'];
$results=$text_results[total_results];
saveToLog($query, $time, $results, $ip, 0);  }
```

3 - المقطع البرمجي المستخدم من أجل الحصول على الصور :

```
if (   $media_only == '1') {
 media_only($query ,$start, $media_only ); }
```

4 - المقطع البرمجي المستخدم للتحقق من عدم تكرار محتوى الصفحة :

```
if ($index_media == '0') {
```

```
$newmd5sum = md5($data['content']); }
if ($md5sum == $newmd5sum) {
 printStandardReport('md5notChanged');
$OKtoIndex = 0; }
 else {
mysqltest();
$result = mysql_query("select link_id from dse_links where md5sum = '$newmd5sum'");
```

5 – المقطع البرمجي المستخدم للتحقق من كون محتوى الصفحة لا يتضمن أي كلمة غير مرغوبة :

```
foreach ($blacklist as $value) {
$met = stripos($data[fulltext], $value);
 if($met)
$found = '1'; }
 if ($found == '1') {
 printStandardReport('matchBlacklist'); }
```

## النتائج و المناقشة:

تم تقييم أداء محرك البحث المصمم اعتماداً على فحص نتائج البحث وتطبيق مقاييس التقييم عليها، وفي هذه الحالة سيكون لدينا أربع فئات من النتائج موضحة في الجدول (11) ينبغي وضعها في الاعتبار وهي :

- مواد ذات صلة بالموضوع تم استرجاعها.
- مواد ذات صلة بالموضوع لم يتم استرجاعها.
- مواد لا صلة لها بالموضوع تم استرجاعها.
- مواد لاصلة لها بالموضوع ولم يتم استرجاعها.

وتعتمد تلك النتائج على بعدين هما:

- الاسترجاع.
- الصلة بالموضوع.

ويتم إعادة النظر إلى الفئات الأربعة وتطبيق معيارين رئيسين عليها لتقييم النتائج وهما: الاستدعاء والدقة، ويضاف إلى هذه المقاييس أيضا مقياسي الحداثة والكلفة. [6]،[8] ،[10] .

**الجدول (11) فئات النتائج**

| | مواد ذات صلة بالموضوع | مواد لا صلة لها بالموضوع |
|---|---|---|
| مواد تم استرجاعها | a | c |
| مواد لم يتم استرجاعها | b | d |

أولاً : **الاستدعاء** recall: هو مقياس لدرجة اكتمال الاسترجاع، حيث يقيس نسبة المواد ذات الصلة بالموضوع التي تم استرجاعها فعلياً من النظام، وذلك بتطبيق المعادلة: R = a / (a+b) .

حيث تبين المعادلة السابقة أن الاستدعاء هو ناتج قسمة المواد ذات الصلة بالموضوع التي تم استرجاعها على مجموع المواد ذات الصلة التي تم استرجاعها مع المواد ذات الصلة التي لم يتم استرجاعها.

ثانياً : **الدقة** precision: هو مقياس لمدى نقاء الاسترجاع، حيث يقيس نسبة المواد المسترجعة ذات الصلة بالموضوع وذلك بتطبيق المعادلة : P=a/(a+c) .

حيث تبين المعادلة السابقة أن الدقة هو ناتج قسمة المواد ذات الصلة بالموضوع التي تم استرجاعها على مجموع المواد ذات الصلة التي تم استرجاعها مع المواد التي لا صلة لها بالموضوع وتم استرجاعها.

ثالثاً :**الحداثة** Novelty: يقصد بالجدة نسبة النتائج الجديدة على المستخدم، أي تلك التي لم يتعرف عليها من قبل، وتأخذ الصيغة التالية :N=n/(n+o) .

حيث n تمثل عدد النتائج الجديدة التي لم يسبق للمستخدم الاطلاع عليها، في حين تمثل o عدد النتائج المعروفة بالنسبة للمستخدم.

رابعاً :**التكلفة** cost: في الحالات التي يكون فيها المستخدم مطالباً بدفع مقابل مادي عن إجراء البحث، فإنه يمكن حساب التكلفة المباشرة مقابل كل نتيجة ذات صلة تم استرجاعها ضمن نتيجة البحث، فعلى سبيل المثال لو أن البحث استرجع 20 نتيجة من بينها 12 نتيجة ذات صلة فعلية بالموضوع وكانت تكلفة البحث 50 ليرة ، فإن تكلفة كل نتيجة تكون 4.16 = 12 / 50 ليرة.

أما لو استرجع البحث 200 نتيجة من بينها 12 نتيجة ذات صلة وكانت تكلفة البحث 150 ليرة ، فإن تكلفة استرجاع النتيجة الواحدة تكون 12.5 = 12 / 150 ليرة. أي أنها ثلاثة أضعاف تكلفة النتيجة في البحث الأول، وبذلك فإن نسبة الدقة والاستدعاء في البحث لها تأثير على التكلفة، فكلما كان البحث أفضل من حيث الاستدعاء والدقة أدى ذلك إلى احتمال نقص تكلفة كل نتيجة ذات صلة بالموضوع يتم استرجاعها.

وعند دراسة العلاقة بين حجم قاعدة البيانات و نتائج تطبيق المقاييس السابقة مع اعتماد تكلفة ثابتة لإجراء البحث تبين أنه عند البحث عن المواضيع المتعلقة بالكلمة computer مع قيام العنكبوت بتعقب روابط خمسة مواقع هي www.cec-co.net و http://www.aya.sy و http://www.9or.cc و http://www.mtn.com.sy و http://www.arabmedmag.com وقيام المفهرس بفهرسة صفحات تلك الروابط ظهرت لدينا النتائج التالية: تضم قاعدة البيانات 16مستند في موضوع الحواسيب، و استُرجع 16 منها عند إجراء البحث، وبذلك فإن نسبة الاستدعاء في هذه النتيجة هي: R=16/(16+0) =1 .

وبفحص النتائج السابقة تبين أن النتائج 16 لها صلة فعلية بموضوع البحث، فإنه يمكن حساب نسبة الدقة على النحو الآتي: P=16/(16+0)=1 .

وبفحص النتائج السابقة تبين أن 7 نتائج منها سبق لنا الاطلاع عليها من قبل، في حين أن 9 نتائج تعد جديدة بالنسبة لنا، وبذلك يمكن احتساب الحداثة على النحو الآتي: N= 9/(9+7) = 0.56

والبحث السابق استرجع 16 نتيجة ذات صلة فعلية بالموضوع و إذا كانت تكلفة البحث 50 ليرة، فإن تكلفة كل نتيجة تكون 3.125 = 16 / 50 ليرة سورية .

وعند البحث عن المواضيع المتعلقة بالكلمة computer مع قيام العنكبوت بتعقب روابط سبعة مواقع هي : http://www.aya.sy و www.cec-co.net و http://en.wikipedia.org و http://www.mtn.com.sy و http://www.9or.cc و http://www.arabmedmag.com و http://www.godssimpleplan.org/gsps-greek.html وقيام المفهرس بفهرسة صفحات تلك الروابط ظهرت لدينا النتائج التالية : تضم قاعدة البيانات 33 مستنداً في موضوع الحواسب، و استُرجع 33 منها عند إجراء البحث، وبذلك فإن نسبة الاستدعاء في هذه النتيجة هي: R=33/(33+0) =1 .

وبفحص النتائج السابقة تبين أن 30 منها فقط لها صلة فعلية بموضوع البحث، في حين 3 نتائج تم استرجاعها لا صلة لها بالموضوع ، فإنه يمكن حساب نسبة الدقة على النحو الآتي: P=30/(30+3)=0.9 .

وبفحص النتائج السابقة تبين أن 16 نتيجة منها سبق لنا الاطلاع عليها من قبل، في حين أن 17 نتيجة تعد جديدة بالنسبة لنا، وبذلك يمكن احتساب الحداثة على النحو الآتي: N= 17/(17+16) = 0.5 .

والبحث السابق استرجع 30 نتيجة ذات صلة فعلية بالموضوع وإذا كانت تكلفة البحث 50 ليرة ، فإن تكلفة كل نتيجة تكون 1.6 = 30 / 50 ليرة سورية.

وعند البحث عن المواضيع المتعلقة بالكلمة computer مع قيام العنكبوت بتعقب روابط تسعة مواقع هي: http://index.arabseyes.com و http://en.wikipedia.org و www.cec-co.net و http://www.aya.sy و http://www.arabmedmag.com و http://www.9or.cc و http://www.mtn.com.sy و http://www.godssimpleplan.org/gsps-greek.html و http://www.runnet.sy وقيام المفهرس بفهرسة صفحات تلك الروابط ظهرت لدينا النتائج التالية : تضم قاعدة البيانات 42 مستنداً في موضوع الحواسب، و استُرجع 42 منها عند إجراء البحث، وبذلك فإن نسبة الاستدعاء في هذه النتيجة هي: R=42/(42+0) =1 .

وبفحص النتائج السابقة تبين أن 39 منها فقط لها صلة فعلية بموضوع البحث، في حين 3 نتائج تم استرجاعها لا صلة لها بالموضوع ، فإنه يمكن حساب نسبة الدقة على النحو الآتي: P=39/(39+3)=0.92 .

و بفحص النتائج السابقة تبين أن 33 نتيجة منها سبق لنا الاطلاع عليها من قبل، في حين أن 9 نتائج تعد جديدة بالنسبة لنا، وبذلك يمكن احتساب الحداثة على النحو الآتي: N= 9/(9+33) = 0.2 .

و البحث السابق استرجع 39 نتيجة ذات صلة فعلية بالموضوع و إذا كانت تكلفة البحث 50 ليرة، فإن تكلفة كل نتيجة تكون 1.28 = 39 / 50 ليرة سورية .

وعند البحث عن المواضيع المتعلقة بالكلمة computer مع قيام العنكبوت بتعقب روابط اثنا عشر موقع هي: http://index.arabseyes.com و http://en.wikipedia.org و www.cec-co.net و http://www.aya.sy و http://www.arabmedmag.com و http://www.9or.cc و http://www.mtn.com.sy و http://www.godssimpleplan.org/gsps-greek.html و http://www.runnet.sy و http://www.syria-news.com و http://www.syriatel.sy و http://www.farisissa.com وقيام المفهرس بفهرسة صفحات تلك الروابط ظهرت لدينا النتائج التالية: تضم قاعدة البيانات 43 مستنداً في موضوع الحواسب، واستُرجع 43 منها عند إجراء البحث، وبذلك فإن نسبة الاستدعاء في هذه النتيجة هي: R=43/(43+0) =1 .

وبفحص النتائج السابقة تبين أن 40 منها فقط لها صلة فعلية بموضوع البحث، في حين 3 نتائج تم استرجاعها لا صلة لها بالموضوع، فإنه يمكن حساب نسبة الدقة على النحو الآتي: P=40/(40+3)=0.93 .

وبفحص النتائج السابقة تبين أن 42 نتيجة منها سبق لنا الاطلاع عليها من قبل، في حين أن نتيجة واحدة فقط تعد جديدة بالنسبة لنا، وبذلك يمكن احتساب الحداثة على النحو الآتي: N= 1/(1+42) = 0.3 .

والبحث السابق استرجع 42 نتيجة من بينها 40 نتيجة ذات صلة فعلية بالموضوع وإذا كانت تكلفة البحث 50 ليرة ، فإن تكلفة كل نتيجة تكون 40 / 50 = 1.25 ليرة سورية.

وعند البحث عن المواضيع المتعلقة بالكلمة computer مع قيام العنكبوت بتعقب روابط عشرين موقع وقيام المفهرس بفهرسة صفحات تلك الروابط ظهرت لدينا النتائج التالية: تضم قاعدة البيانات 45 مستنداً في موضوع الحواسب، واستُرجع 45 منها عند إجراء البحث، وبذلك فإن نسبة الاستدعاء هي R=45/(45+0)=1 .

وبفحص النتائج السابقة تبين أن 40 منها فقط لها صلة فعلية بموضوع البحث، في حين أن 5 نتائج تم استرجاعها لا صلة لها بالموضوع ، فإنه يمكن حساب نسبة الدقة على النحو الآتي: P=40/(40+5)=0.88 .

و بفحص النتائج السابقة تبين أن 43 نتيجة منها سبق لها الاطلاع عليها من قبل، في حين أن 2 نتيجة فقط تعد جديدة بالنسبة لنا، وبذلك يمكن احتساب الحداثة على النحو الآتي: N= 2/(2+43) = 0.04 .

و البحث السابق استرجع 35 نتيجة ذات صلة فعلية بالموضوع وإذا كانت تكلفة البحث 50 ليرة ، فإن تكلفة كل نتيجة تكون 40 / 50 = 1.25 ليرة سورية.

وعند البحث عن المواضيع المتعلقة بالكلمة computer مع قيام العنكبوت بتعقب روابط خمس وثلاثين موقع و قيام المفهرس بفهرسة صفحات تلك الروابط ظهرت لدينا النتائج التالية: تضم قاعدة البيانات 50 مستنداً في موضوع الحواسب، واستُرجع 50 منها عند إجراء البحث، وبذلك فإن نسبة الاستدعاء هي:R=50/(50+0)=1.

وبفحص النتائج السابقة تبين أن 44 منها فقط لها صلة فعلية بموضوع البحث، في حين أن 6 نتائج تم استرجاعها لا صلة لها بالموضوع، فإنه يمكن حساب نسبة الدقة على النحو الآتي: P=44/(44+6)=0.88 .

و بفحص النتائج السابقة تبين أن 45 نتيجة منها سبق لنا الاطلاع عليها من قبل، في حين أن 5 نتائج فقط تعد جديدة بالنسبة لنا، وبذلك يمكن احتساب الحداثة على النحو الآتي: N= 5/(5+45) = 0.1 .

و البحث السابق استرجع 38 نتيجة ذات صلة فعلية بالموضوع وإذا كانت تكلفة البحث 50 ليرة ، فإن تكلفة كل نتيجة تكون 44 / 50 = 1.25 ليرة سورية.

وعند البحث عن المواضيع المتعلقة بالكلمة computer مع قيام العنكبوت بتعقب روابط خمسين موقع و قيام المفهرس بفهرسة صفحات تلك الروابط ظهرت لدينا النتائج التالية : تضم قاعدة البيانات 60 مستنداً في موضوع الحواسب، و استُرجع 60 منها عند إجراء البحث، وبذلك فإن نسبة الاستدعاء هي: R=60/(60+0)=1 .

وبفحص النتائج السابقة تبين أن 49 منها فقط لها صلة فعلية بموضوع البحث، في حين 11 نتيجة تم استرجاعها لا صلة لها بالموضوع، فإنه يمكن حساب نسبة الدقة على النحو الآتي : P=49/(49+11)=0.81 .

وبفحص النتائج السابقة تبين أن 50 نتيجة منها سبق لنا الاطلاع عليها من قبل، في حين أن 10 نتائج تعد جديدة بالنسبة لنا، وبذلك يمكن احتساب الحداثة على النحو الآتي: N= 10/(10+50) = 0.16 .

و البحث السابق استرجع 40 نتيجة ذات صلة فعلية بالموضوع و إذا كانت تكلفة البحث 50 ليرة ، فإن تكلفة كل نتيجة تكون 49 / 50 = 1.02 ليرة سورية.

وعند البحث عن المواضيع المتعلقة بالكلمة computer مع قيام العنكبوت بتعقب روابط خمس و سبعين موقع و قيام المفهرس بفهرسة صفحات تلك الروابط ظهرت لدينا النتائج التالية : تضم قاعدة البيانات 88 مستنداً في موضوع الحواسب، واستُرجع 88 منها عند إجراء البحث، وبذلك فإن نسبة الاستدعاء هي: R=88/(88+0) =1. وبفحص النتائج السابقة تبين أن 66 منها لها صلة فعلية بموضوع البحث، في حين أن 22 نتيجة تم استرجاعها لا صلة لها بالموضوع، فإنه يمكن حساب نسبة الدقة على النحو الآتي: P=66/(66+22)=0.75 .

و بفحص النتائج السابقة تبين أن 60 نتيجة منها سبق لنا الاطلاع عليها من قبل، في حين أن 28 نتيجة تعد جديدة بالنسبة لنا، وبذلك يمكن احتساب الحداثة على النحو الآتي: N= 28/(28+60) = 0.31 .

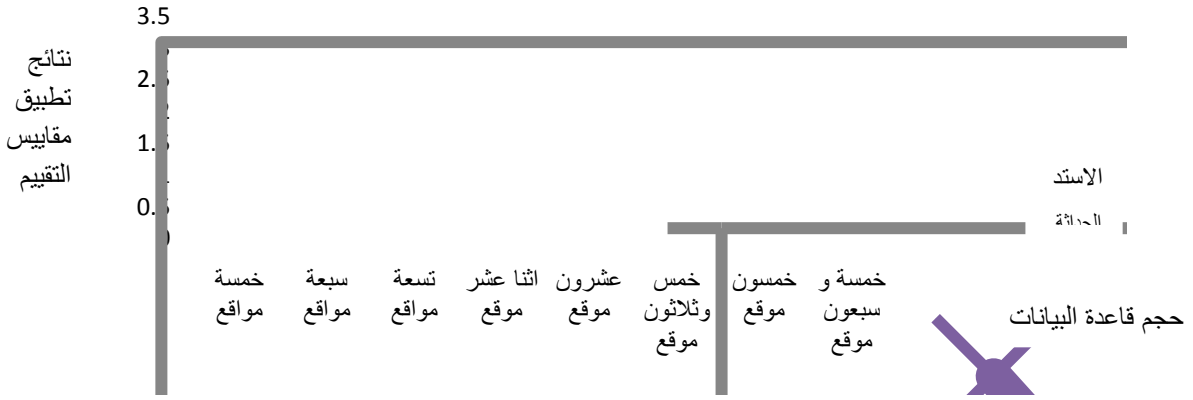
و البحث السابق استرجع 68 نتيجة ذات صلة فعلية بالموضوع و إذا كانت تكلفة البحث 50 ليرة ، فإن تكلفة كل نتيجة تكون 0.75 = 66 / 50 ليرة سورية.

و يوضح الشكل (6) المخطط البياني لتقييم أداء محرك البحث المقترح حيث نلاحظ أن نسبة الاستدعاء ثابتة و تساوي 1، كون محرك البحث يطابق الطلب كحروف لغوية وهذا يؤدي لاستدعاء كامل النتائج التي تطابق كلمة البحث وبالتالي نسبة الاستدعاء لن تتغير و تبقى ثابتة مساوية 1.

كما نلاحظ أن نسبة الدقة تنخفض بشكل بسيط مع زيادة حجم قاعدة البيانات حيث أنه مع ازدياد حجم قاعدة البيانات تزداد النتائج التي لا صلة لها بموضوع البحث وبالتالي تنخفض نسبة الدقة.

و نلاحظ أيضاً أن نسبة الحداثة تنخفض بشكل صغير مع ازدياد حجم قاعدة البيانات حيث إنه مع ازدياد حجم قاعدة البيانات تنخفض نسبة النتائج الجديدة التي لم يسبق للمستخدم الاطلاع عليها وبالتالي تنخفض نسبة الحداثة.

كما نلاحظ أن التكلفة تنخفض بشكل كبير مع زيادة حجم قاعدة البيانات حيث إنه مع زيادة حجم قاعدة البيانات يزداد عدد النتائج التي يتم استرجاعها و بالتالي تنخفض تكلفة كل نتيجة يتم استرجاعها.




**الشكل (6) المخطط البياني لتقييم أداء محرك البحث المقترح**

وعند البحث عن المواضيع الطبية بإدخال الكلمة teeth مع قيام العنكبوت بتعقب روابط 4 مواقع وقيام المفهرس بفهرسة صفحات تلك الروابط ظهرت لدينا النتائج التالية: تضم قاعدة البيانات 11 مستند في المواضيع الطبية، و استُرجع 11 منها عند إجراء البحث، وبذلك فإن نسبة الاستدعاء هي: R=11/(11+0) =1.

وبفحص النتائج السابقة تبين أن النتائج 10 لها صلة فعلية بموضوع البحث، في حين نتيجة فقط تم استرجاعها لا صلة لها بالموضوع، فإنه يمكن حساب نسبة الدقة على النحو الآتي: P=10/(10+1)=0.9.

وبفحص النتائج السابقة تبين أن 3 نتائج منها سبق لنا الاطلاع عليها من قبل، في حين أن 8 نتائج تعد جديدة بالنسبة لنا، وبذلك يمكن حساب الحداثة على النحو الآتي: 0.73 = (8+3)/8 =N.

والبحث السابق استرجع 11 نتيجة ذات صلة فعلية بالموضوع وإذا كانت تكلفة البحث 50 ليرة، فإن تكلفة كل نتيجة تكون 4.55 = 11 / 50 ليرة سورية.

وعند البحث عن المواضيع المتعلقة بالكلمة teeth مع قيام العنكبوت بتعقب روابط 6 مواقع وقيام المفهرس بفهرسة صفحات تلك الروابط ظهرت لدينا النتائج التالية : تضم قاعدة البيانات 25 مستنداً في موضوع الأسنان، و استُرجع 25 منها عند إجراء البحث، وبذلك فإن نسبة الاستدعاء في هذه النتيجة هي: 1= (25+0)/25=R .

وبفحص النتائج السابقة تبين أن 21 منها فقط لها صلة فعلية بموضوع البحث، في حين 4 نتائج تم استرجاعها لا صلة لها بالموضوع ، فإنه يمكن حساب نسبة الدقة على النحو الآتي: 0.84=(21+4)/21=P .

وبفحص النتائج السابقة تبين أن 15 نتيجة منها سبق لنا الاطلاع عليها من قبل، في حين أن 10 نتائج تعد جديدة بالنسبة لنا، وبذلك يمكن احتساب الحداثة على النحو الآتي: 0.4 = (10+15)/10 =N .

والبحث السابق استرجع 21 نتيجة ذات صلة فعلية بالموضوع وإذا كانت تكلفة البحث 50 ليرة ، فإن تكلفة كل نتيجة تكون 2.38 = 21 / 50 ليرة سورية.

وعند البحث عن المواضيع المتعلقة بالكلمة teeth مع قيام العنكبوت بتعقب روابط 9 مواقع وقيام المفهرس بفهرسة صفحات تلك الروابط ظهرت لدينا النتائج التالية : تضم قاعدة البيانات 30 مستنداً في موضوع الأسنان، و استُرجع 30 عند إجراء البحث، وبذلك فإن نسبة الاستدعاء في هذه النتيجة هي: 1= (30+0)/30=R .

وبفحص النتائج السابقة تبين أن 25 منها فقط لها صلة فعلية بموضوع البحث، في حين 5 نتائج تم استرجاعها لا صلة لها بالموضوع ، فإنه يمكن حساب نسبة الدقة على النحو الآتي: 0.83=(25+5)/25=P .

وبفحص النتائج السابقة تبين أن 20 نتيجة منها سبق لنا الاطلاع عليها من قبل، في حين أن 10 نتائج تعد جديدة بالنسبة لنا، وبذلك يمكن احتساب الحداثة على النحو الآتي: 0.33 = (10+20)/10 =N .

و البحث السابق استرجع 25 نتيجة ذات صلة فعلية بالموضوع و إذا كانت تكلفة البحث 50 ليرة، فإن تكلفة كل نتيجة تكون 2 = 25 / 50 ليرة سورية.

وعند البحث عن المواضيع المتعلقة بالكلمة teeth مع قيام العنكبوت بتعقب روابط 15 موقع وقيام المفهرس بفهرسة صفحات تلك الروابط ظهرت لدينا النتائج التالية: تضم قاعدة البيانات 33 مستنداً في موضوع الأسنان، واستُرجع 33 منها عند إجراء البحث، وبذلك فإن نسبة الاستدعاء في هذه النتيجة هي: 1= (33+0)/33=R .

وبفحص النتائج السابقة تبين أن 27 منها فقط لها صلة فعلية بموضوع البحث، في حين 6 نتائج تم استرجاعها لا صلة لها بالموضوع، فإنه يمكن حساب نسبة الدقة على النحو الآتي: 0.81=(27+6)/27=P .

وبفحص النتائج السابقة تبين أن 26 نتيجة منها سبق لنا الاطلاع عليها من قبل، في حين أن 7 نتائج تعد جديدة بالنسبة لنا، وبذلك يمكن احتساب الحداثة على النحو الآتي: 0.21 = (7+26)/7 =N .

والبحث السابق استرجع 27 نتيجة ذات صلة فعلية بالموضوع وإذا كانت تكلفة البحث 50 ليرة ، فإن تكلفة كل نتيجة تكون 1.85 = 27 / 50 ليرة سورية.

وعند البحث عن المواضيع المتعلقة بالكلمة teeth مع قيام العنكبوت بتعقب روابط 35 موقع وقيام المفهرس بفهرسة صفحات تلك الروابط ظهرت لدينا النتائج التالية : تضم قاعدة البيانات 35 مستنداً في موضوع الأسنان، واستُرجع 35 منها عند إجراء البحث، وبذلك فإن نسبة الاستدعاء في هذه النتيجة هي: R=35/(35+0) =1 .

وبفحص النتائج السابقة تبين أن 24 منها فقط لها صلة فعلية بموضوع البحث، في حين أن 11 نتيجة تم استرجاعها لا صلة لها بالموضوع، فإنه يمكن حساب نسبة الدقة على النحو الآتي: P=24/(24+11)=0.68 .

و بفحص النتائج السابقة تبين أن 34 نتيجة منها سبق لنا الاطلاع عليها من قبل، في حين أن نتيجة فقط تعد جديدة بالنسبة لنا، وبذلك يمكن احتساب الحداثة على النحو الآتي: N= 1/(1+34) = 0.02 .

و البحث السابق استرجع 24 نتيجة ذات صلة فعلية بالموضوع وإذا كانت تكلفة البحث 50 ليرة ، فإن تكلفة كل نتيجة تكون 50 / 24 = 2.08 ليرة سورية.

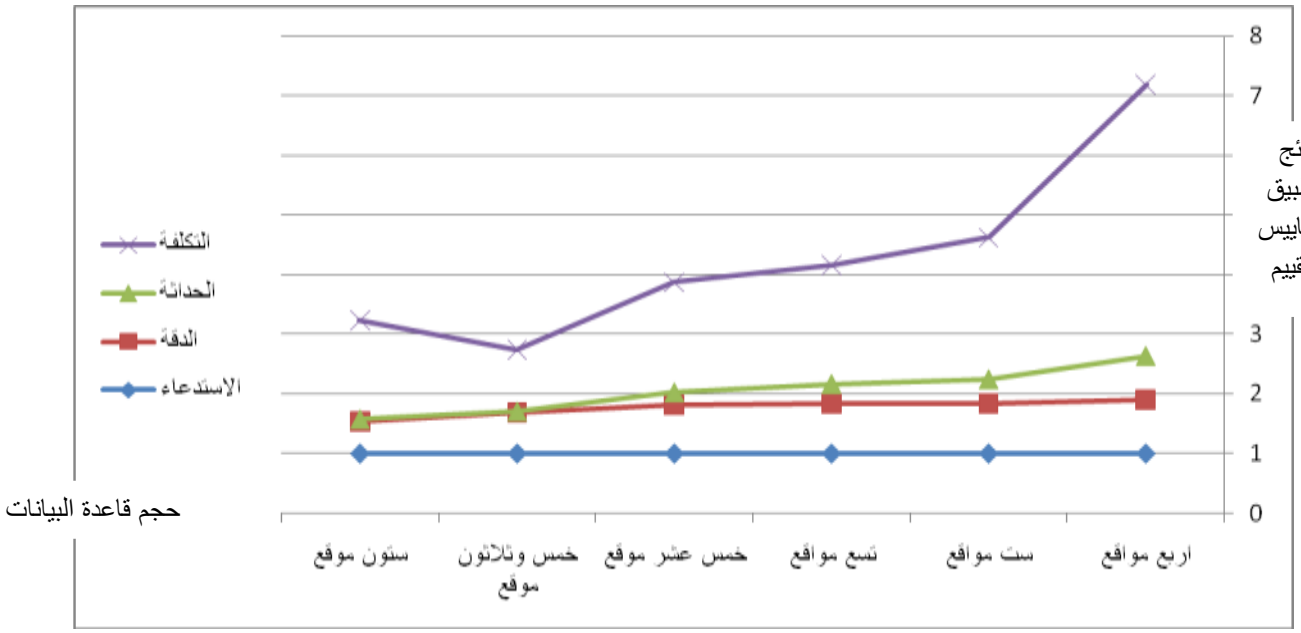
وعند البحث عن المواضيع المتعلقة بالكلمة teeth مع قيام العنكبوت بتعقب روابط 60 موقع و قيام المفهرس بفهرسة صفحات تلك الروابط ظهرت لدينا النتائج التالية : تضم قاعدة البيانات 55 مستنداً في موضوع الأسنان، و استُرجع 55 منها عند إجراء البحث، وبذلك فإن نسبة الاستدعاء في هذه النتيجة هي: R=55/(55+0) =1 .

وبفحص النتائج السابقة تبين أن 30 منها فقط لها صلة فعلية بموضوع البحث، في حين 25 نتائج تم استرجاعها لا صلة لها بالموضوع، فإنه يمكن حساب نسبة الدقة على النحو الآتي : P=30/(30+25)=0.54 .

وبفحص النتائج السابقة تبين أن 53 نتيجة منها سبق لنا الاطلاع عليها من قبل، في حين أن 2 نتيجة تعد جديدة بالنسبة لنا، وبذلك يمكن احتساب الحداثة على النحو الآتي: N= 2/(2+53) = 0.03 .

و البحث السابق استرجع 30 نتيجة ذات صلة فعلية بالموضوع و إذا كانت تكلفة البحث 50 ليرة ، فإن تكلفة كل نتيجة تكون 50 / 30 = 1.66 ليرة سورية.

و يوضح الشكل (7) المخطط البياني لتقييم أداء محرك البحث المقترح.




**الشكل (7) المخطط البياني لتقييم أداء محرك البحث المقترح**

و عند البحث عن مواضيع غير مرغوبة كالمواضيع الفنية و ذلك بإدخال كلمة songs مثلا، فلا تظهر نتائج و ذلك لأن محرك البحث المقترح يتميز بقائمة حجب تحتوي الكلمات و المواضيع غير المرغوبة.

كما أنه يحتوي أيضا على قائمة حجب تضم الروابط غير المرغوبة فإذا صادف المفهرس الرابط التالي مثلا و هو رابط للأغاني فيقوم بعدم فهرسة هذا الرابط والانتقال إلى الرابط التالي لاختباره.

## الاستنتاجات والتوصيات:

بُنيت نواة محرك بحث ويب يتحكم بالبحث و الفهرسة، يمكن أن يكون منصة باتجاه بناء محرك بحث منافس لمحركات البحث العالمية الموجودة. وقد تركزت الجهود للاهتمام بجميع التفاصيل و الخيارات لبناء قواعد البيانات و مكونات المحرك الأخرى. حيث أن محرك البحث المقترح يمكن من خلاله البحث عن النتائج النصية التي تطابق كلمة البحث كحروف لغوية دون الغوص بالمعنى و يتميز بقدرته على حجب الروابط و كلمات البحث غير المرغوبة، و لاحظنا من نتائج تطبيق مقاييس التقييم على نتائج البحث أن شمولية الاسترجاع ثابتة مهما تغير حجم قاعدة البيانات ولكن مع زيادة حجم قاعدة البيانات ينخفض نقاء الاسترجاع قليلاً و تنخفض نسبة النتائج الجديدة التي لم يسبق للمستخدم الاطلاع عليها بشكل بسيط كما تنخفض تكلفة كل نتيجة تم استرجاعها بشكل كبير، وبالتالي فان اختيار حجم قاعدة بيانات ملائم وتحديثها بشكل مستمر يمكننا من زيادة معياري النقاء و الحداثة و تخفيض معيار الكلفة و زمن البحث المستغرق للحصول على النتائج مما يجعله ذو فعالية أكبر.

## المراجع:

[1] GREHAN,M. *How Search Engines Work*, Incisive Media , New York,2002,275.

[2] KENT,P. *Search Engine Optimization*.5th.ed., John Wiley & Sons , London ,2012 ,456.

[3] MENG,W. *Metasearch Engines*, Binghamton University, New York, 2008,302.

[4] DING,C.H. ; BUYYA,R. *Guided Google: A Meta Search Engine and its Implementation using the Google Distributed Web Services*,University of Melbourne, Australia,2006,455.

[5] OLSTON,C.; NAJORK,M. *Web Crawling*. Foundations and Trends in Information Retrieval, Vol. 4, No. 3 ,2010,175–246.

[6] CROFT,B.; METZLER,D. ;STROHMAN,T. *Search Engines: Information Retrieval in Practice*, Addison-Wesley,2010,520.

[7] HERSH,W. *information retrieval*.3th.ed., Springer, German ,2010,486.

[8] MANNING,C.; RAGHAVAN,P.; SCHÜTZE,H., An *Introduction to Information Retrieval*, Cambridge UP , England, ,2009,544.

[9] LEVENE,M. *An Introduction to Search Engines and Web* Navigation, Wiley, London,2010,334.

[10] BAMEFLIH,F. ,*The basics of electronic Information retrieval systems*.1st.ed.,KFNL , Riyadh, 2006, 329.