

Building a Search Engine for the Personal Computer and the Local Network

Fares Adel Dawood

Khaled Hussein

Mohamed Majed

College of Computer Science and Mathematics
University of Mosul, Mosul, Iraq

Received on: 24/10/2012

Accepted on: 30/01/2013

ABSTRACT

Computer users need to advanced methods for accessing stored information with at high speed and accurate information retrieval of from among the millions of files stored on a personal computer when they request. This research aims to design and construct a search engine works on personal computer or Intranet to retrieve required files, using natural language dialogue interaction with the user and the computer, rely on key words as an essential element semantic system available on operating systems that support only path and file name in the search.

The search engine programme (**Search Engine for PC and Intranet - SPCI**) covering the most important file types that the user interacts on an ongoing basis and most text files where the text file can be retrieved only by mentioning a word or phrase within the text.

The (SPCI) programmer has been developed under Microsoft visual C# version 3.0
Keywords: Search Engine, Intranet, SPCI.

بناء محرك بحث للحاسوب الشخصي والشبكة المحلية

محمد ماجد

خالد حسين

فارس عادل داؤد

كلية علوم الحاسوب والرياضيات، جامعة الموصل

تاريخ قبول البحث: 2013/01/30

تاريخ استلام البحث: 2012/10/24

المخلص

يحتاج مستخدمو الحاسوب إلى طرائق مطورة للوصول إلى المعلومات المخزونة بسرعة عالية ودقة في استرجاع المعلومات من بين ملايين الملفات المخزونة في الحاسوب الشخصي عند الطلب. يهدف هذا البحث إلى تصميم وبناء محرك بحث يعمل على الحاسوب الشخصي أو شبكة الانترنت (Intranet) لاسترجاع ملفات مطلوبة من خلال لغة التحوار الطبيعية مع المستخدم تعتمد على المفاتيح الأساسية (Keywords) عنصراً أساسياً في الاسترجاع على العكس من الأنظمة الدلالية التقليدية المتوفرة في أنظمة التشغيل والتي تعتمد فقط على اسم الملف ومساره في البحث.

يشمل العمل في برنامج محرك البحث (Search engine for pc and Intranet - SPCI) شمول أهم أنواع الملفات التي يتعامل معها المستخدمون بصورة مستمرة ومن أشهرها الملفات النصية حيث يمكن استرجاع ملف نصي فقط بذكر كلمة أو عبارة موجودة داخل النص.

لقد تم تطوير برنامج (SPCI) بالاعتماد على لغة C# فيجوال بالإصدار 3.0، يوفر برنامج SPCI

أمكانية البحث عن ملفات بامتدادات (Excel, Word, Pdf, Txt, Html).

الكلمات المفتاحية: محرك بحث، الشبكة المحلية، SPCI.

1- المقدمة

محرك البحث (الباحوث) هو عبارة عن برامج مجانية متوفرة خلال مواقع خاصة على الانترنت تتيح للمستخدم البحث عن معلومات أشخاص أو ملفات ضمن مصادر الانترنت المختلفة، وتعتمد هذه المحركات على الفهرسة الآلية برصد التعابير والمفردات والكلمات المفتاحية الواردة في المعلومات المنشورة في مصادر الانترنت ويتم البحث فيها باستخدام الكلمات المفتاحية (keywords) ويتميز محركات البحث باحتوائها على معلومات أكثر من تلك المعلومات التي تحتويها أدلة البحث (directories) [1].

وتعرض النتائج على شكل قائمة بعناوين المستندات التي توافق الطلب. يرفق بالعناوين في الغالب مختصر عن المستند المشار إليه أو جزء من محتواه. عناصر قائمة البحث ترتب على حسب معايير خاصة (وقد تختلف من محرك لأخر) من أهمها مدى موافقة كل عنصر للطلب [1]. وعند الحديث عن محركات البحث فغالبا ما يقصد محركات البحث على شبكة الإنترنت ومحركات الويب بالخصوص. محركات البحث في الويب تبحث عن المعلومات على الشبكة العنكبوتية العالمية، ومنها يستعمل على نطاق ضيق يشمل البحث داخل الشبكات المحلية للمؤسسات أي إنترانت (Intranet) [9]. أما محركات البحث الشخصية فتبحث في الحواسيب الشخصية الفردية [2].

والهدف من البحث هو بناء وتصميم برنامج (SPCI) يعمل على زيادة في سرعة ودقة استرجاع البيانات المخزونة في الحواسيب الشخصية المعزولة أو التي تكون متصلة على شبكة محلية (Intranet) في مؤسسات مختلفة.

2- الأعمال السابقة

قبل مجيء الويب (وهي الشبكة العامة قابلة النفاذ للجميع)، كانت هناك محركات بحث لبروتوكولات أو أنظمة قديمة، مثل محرك بحث آركي Archie لمواقع إف تي بي المجهولة anonymous FTP ومحرك بحث فيرونيكا Veronica لبروتوكول جوفر Gopher.

وبعد سنتين من ظهور جوفر بشكله المطور، بدأت خدمات ويب العملاقة في الظهور، والتي تسمح باستخدام الوصلات التشعبية والبحث النصي والاستعراض البياني [6].

في عام 1993 كان أول محرك بحث ويب هو واندكس Wandex، القائم على فهرس التجول في جميع أنحاء العالم عبر الانترنت World Wide Web Wanderer وهو زاحف عنكبوتي web crawler تم تطويره في معهد ماساشوستس للتكنولوجيا MIT.

وأیضا في عام 1993 ظهر محرك بحث آخر هو آليويب Aliweb، ولا زال ساريا اليوم. وفي عام 1994 كان ظهور أول محرك بحث قائم على الزاحف العنكبوتي للنصوص الكاملة فكان وبكراولر WebCrawler تم تطويره في جامعة واشنطن .

وفي عام 2004 ظهر لايكوس Lycos (الذي بدأ في جامعة كارنيجي ميلون Carnegie Mellon University) وصار مشروعا تجاريا كبيرا.

وأیضا في عام 2004 ظهرت العديد من محركات البحث وتنافست على الرواج وكان من ضمنها إكساييت Excite، إنفوسيك Infoseek، إنكتومي Inktomi، نورثرن لايت Northern Light، وألتايفاستا Alta

Vista. وفي بعض الحالات تنافست مع الأدلة ذات الشعبية مثل ياهو Yahoo. فيما بعد، أدمجت الأدلة مع تقنية محرك البحث من أجل أداء أكبر للوظائف [6].

3- محركات البحث المتداولة

3-1 محركات البحث على الإنترنت

في الوقت الحاضر تعد مواقع البحث على شبكة الإنترنت من أكثر المواقع التي تلاقي إقبالا من مستخدمي الشبكة إذ يستفيد منها ملايين المستخدمين لمساعدتهم في الوصول إلى عناوين المواقع التي تحتوي على المعلومات التي يحتاجونها، فمواقع محركات البحث هي أفضل وأقصر طريق للوصول إلي المعلومات التي يحتاجها المستخدمون كما أن هذه المواقع مجانية ويستفيد منها المستخدم دون دفع أي مقابل [7]. وفيما يلي نعرض بعضاً من أهم وأكبر مواقع محركات البحث التي يمكننا الاستفادة منها على الإنترنت [7].

- موقع البحث (Google).
- ياهو (Yahoo!)
- بينج (Bing)
- ألتافيستا (AltaVista)
- إم إس إن ميكروسوفت (Msn)
- لايف سيرش (LiveSearch)

3-2 محركات البحث على الحاسوب الشخصي

إن محركات البحث تنتج من قبل شركات البرمجيات العالمية والربحية، والتي دوماً ما تكون قيد التطوير مع مرور الزمن سعياً لتحقيق أفضل النتائج من ناحية الدقة والسرعة. على هذا الأساس قمنا بجمع معلومات حول محركات البحث الموجودة والمتداولة بين المستخدمين، ومن خلال دراستنا لتلك البرامج المنتجة من قبل الشركات العالمية والربحية، وجدنا البعض منها يقوم بالبحث عن ملفات نصية وصورية، والبعض الآخر يختص بنوع محدد من الملفات، كأن تكون صورية أو نصية، بعض الأنواع تفتقد إلى السرعة في الحصول على تلك النتائج، والأخرى تمتاز بالسرعة ولكنها تبث ضمن نطاق ضيق، كأن يكون على نص أو صورة، ومن خلال دراستنا أيضاً وجدنا بعض محركات البحث تمتاز بالكفاءة والسرعة في الحصول على نتائج سريعة، ولكن افتقدت إلى أمر مهم جداً، وهو إن عملية البحث تجري فقط داخل الملف النصي، والبعض الآخر لا يتمكن من الوصول إلى مسار المجلدات الطويل، وهذا الأمر يعد نقطة ضعف في ذلك المحرك، فالمستخدم لا يرغب ويطلب السرعة والكفاءة لوحدها، وإنما يرغب بالحصول على نتائج صحيحة بالدرجة الأولى. ومع استمرار بحثنا ودراساتنا حول تلك المحركات، استطعنا تجنب بعض المشكلات الأساسية التي ظهرت معهم.

- Windows Search
- Hulbee Desktop search
- Copernic Desktop Search
- Exalead Desktop search

- Docfetcher search
- VoidTools search
- Insight Desktop Search
- FileSeek search
- Ultra Search

4- أجزاء محرك البحث

- ويتألف محرك البحث من ثلاثة أجزاء رئيسية هي [3],[10].
- برنامج العنكبوت (Spider Program). يقوم بالإبحار عبر الانترنت وبصفة دورية لتتبع صفحات الويب للإطلاع على محتوياتها وتسجيل بياناتها من عناوين وكلمات مفتاحيه.
- برنامج المُفهرِس (Indexer Program).
- وظيفته فهرست المعلومات والنصوص التي حصل عليها من المستكشف باستخدام المعايير والتقنيات المبينة على نظام البحث في النصوص الكاملة مثل معايير الكلمات الأكثر تكراراً من غيرها وأدراجها ضمن قاعدة بيانات (Index database) الخاصة به.
- برنامج محرك البحث (Search Program).
- وهو الواجهة التخاطبية بين المستخدم ومحرك البحث إذ يقوم باستقبال الكلمات المفتاحية التي يكتبها المستخدم في مربع البحث (search box) للبحث عنها ضمن مصادر الانترنت المختلفة.

5- خوارزميات البحث

- هنالك خوارزميات متعددة في استرجاع البيانات المخزونة وهذه الخوارزميات تختلف في طريقة الوصول إلى البيانات وكذلك في السرعة لاسترجاع المعلومات وفيما يلي الخوارزميات [9].
- خوارزمية Naive Searching Algorithms(also known as Brute-Force
- خوارزمية Knuth - Morris - Pratt String Matching (KMP)
- خوارزمية Boyer-Moore Algorithm
- خوارزمية Depth First Search

6- طريقة عمل محركات البحث في الحاسوب الشخصي

تعمل محركات البحث عن طريق تخزين معلومات عن عدد كبير من الملفات النصية والصورية وبقية الأنواع الأخرى المعرفة لدى نظام التشغيل بالنسبة للحاسوب، والتي تستعيدها من خلال برنامج الفهرسة التي سبق وان قام بفهرسة جميع تلك الملفات، وتستعاد هذه الملفات بإحدى طرائق استرجاع المعلومات، بعد ذلك يجري تحليل كل ملف لتحديد كيف ينبغي فهرستها (على سبيل المثال، تستخلص الكلمات من العناوين، رؤوس الموضوعات)، وتخزن البيانات عن الملفات في مكدس فهرسي للاستخدام في عمليات البحث طلباً للمعلومات لاحقاً [1].

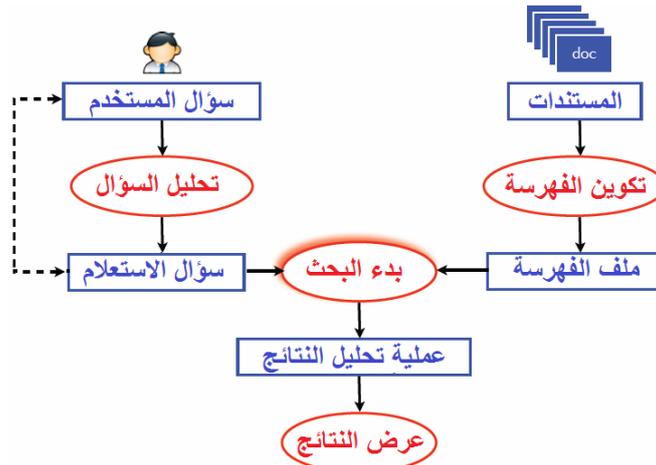
بعض محركات البحث، تعمل على أساس ملف فهرسة، على سبيل المثال محرك البحث الخاص بنظام الويندوز، والأخرى تعتمد على طرائق خاصة لم نستطع اكتشافها لان الشركات المنتجة تحتفظ بالحقوق لها فقط. أما بالنسبة لمحرك البحث الخاص ببحثنا، فإنه يخزن مواقع تلك الملفات في الحاسوب، أي امتدادها وموقع خزنها على

القرص الصلب مع بعض المعلومات عنها التي تساعد وتسرع في عملية البحث، وحينما يتم البحث عن ملف محدد، فإنه يسترجع مواقع الملفات المخزونة في قاعدة البيانات الخاصة به، ومن ثم القيام بعملية تحليل ومطابقة سؤال الاستعلام الذي ادخله المستخدم "الباحث" وإعطاء النتائج المطلوب فقط.

عندما يوجه المستخدم محرك البحث وتجري عملية البحث طلباً لمعلومات معينة، كما هو السائد وذلك بإعطاء كلمات مفتاحية، يفتش المحرك في الفهرس ويقدم قائمة النتائج الأفضل توافقاً تبعاً لمعاييرها، في المعتاد مع ملخص قصير يحتوي على عنوان الوثيقة وأحياناً أجزاء من النص [3].

ويعتمد مدى فائدة محرك بحث على مدى صلة النتائج التي يرد بها، فبينما قد تكون هناك ملايين الملفات التي تحتوي على كلمة أو عبارة محددة، قد تكون بعض أو ثقل صلة، أو أروج، أو معتمدة أكثر من غيرها [3].

معظم محركات البحث توظف أساليب لوضع مراتب النتائج لتقديم أفضل النتائج، أولاً الكيفية التي يقرر بها محرك بحث أي الملفات هي الأفضل توافقاً، وما هي الهيئة الذي يجب أن تظهر بها النتائج، تختلف بشكل شاسع من محرك لآخر، والأساليب أيضاً تتغير عبر الزمن بتغير استخدام تقنية البحث وتقنيات جديدة متطورة [3]. والمخطط (1) يوضح أهم المحاور الأساسية لعمل محرك البحث بشكل مبسط.



الشكل (1). المحاور الأساسية لعمل محرك البحث

1-6 الاستعلام

الاستعلام هو سؤال عن البيانات المخزونة في فهرسة محرك البحث، للبحث عن ملف محدد أو مجموعة من الملفات، إذ إن برنامج الاستعلام قد يجمع جملة من البيانات التي تكون متوفرة في ملف أو أكثر، اعتماداً على الصيغة المكتوبة في صندوق البحث "Search Box".

2-6 الفهرسة

الفهرسة هي ترتيب نتائج الاستعلام في شكل متناسب ومتتالي ليسهل عملية البحث مع مراعاة النتائج فيراعي محرك البحث الترتيب على أن يشمل في نتائجه الموقع التي يوجد به كلمات الاستعلام بكثرة بأدراجها في أول النتائج المدرجة للباحث بالإضافة إلى أن هناك بعض المحركات تعتمد على بعض الخوارزميات الأخرى في تخزين ملفات إضافية يتطلبها محرك البحث، والغرض من الفهرسة هو تحسين السرعة والأداء في العثور على الوثائق ذات الصلة لاستعلام البحث. وبدون فهرسة، فإن محرك البحث يفحص كل وثيقة في الحاسوب، الأمر

الذي يتطلب وقتاً طويلاً وقدرة حاسوبية عالية. على سبيل المثال، يمكن للمؤشر أن يخزن أكثر من 10000 من المستندات الكبيرة ولكن هذه العملية تستغرق ساعات طويلة، ولكن في عملية الاسترجاع لا تستغرق الفترة الزمنية نفسها التي استغرقتها في عملية تكوين الفهرسة [4].

ولكننا كثيراً أثناء بحثنا باللغة العربية نجد انه تم إدراج نتائج بشكل خاطئ أو ابتعاد محرك البحث عما نحصده مع كتابتك لعبارة الاستعلام بشكل جيد هذه كانت مشكلة محركات البحث ولعلنا أحسنا بتأخر ظهور محركات بحث عربية ولكن يرجع السبب في التقنيات المعقدة التي يحتاجها البحث باللغة العربية. إذ تختلف طبيعة اللغة العربية عن الإنجليزية، فاللغة العربية لغة صَرفية (morphological)، في حين أن الإنجليزية لغة لصقية (affixational). ومن هنا كان لا بد للشركات التي تطرح محركات بحث عربية قوية أن تمتلك التقنيات اللازمة لمعالجة اللغة العربية آلياً [4].

3-6 تحليل صيغة الملفات

إذا كان محرك البحث يدعم صيغ متعددة من الملفات، فلا بد من إعداد وثائق لكل الأنواع التي يدعمها المحرك، ولكن التحدي هو أن العديد من الملفات تحتوي على معلومات الملف إضافة إلى المضمون النصي [5]. تحليل الصيغة (أو كما يعرف تحليل الشكل) هو تحديد كيفية التعامل مع محتوى الملف، وهو جزء لا يتجزأ من ضمن العمليات التي تسيطر على عملية البحث والتحليل للملفات ذات الأنواع المختلفة، التي يتم بها تقديم هذه الوثيقة بشكل واضح لجزء البحث في المحرك. ويشار أيضاً إلى الأمور التالية على أنها جزء من مرحلة تحليل الشكل، على سبيل المثال تجريد العلامات والتنسيقات (كما في صيغة صفحات الانترنت)، وتطبيع وإعداد النص، ويزيد من تعقيد التحدي المتمثل في تحليل شكل اختلاف صيغ الملفات، وبما أن صيغ الملفات كثيرة جداً، فإن ذلك يتطلب توفير عمليات متعددة من تحليل الشكل لكل تنسيق [5].

هنالك بعض الصيغ موثقة توثيقاً جيداً ومدعومة في اغلب محركات البحث هي:-

- صيغ الكتب الالكترونية (PDF).
- مايكروسوفت وورد (DOC).
- مايكروسوفت اكسل (EXL).
- ملفات نصية ASCII (نص الوثيقة بدون تنسيق كمبيوتر معين للقراءة (TXT)) [5].

7- تطوير برنامج (SPCI)

يمتاز برنامج Search Engine For Personal Computer And Intranet (SPCI) بسرعة استرجاع المعلومات بكفاءة عالية ودقة على الحواسيب الشخصية والحواسيب المتصلة بشبكة الانترنت والمخطط (3 إلى 7) يوضح عمل أجزاء البرنامج، لقد تم تطوير البرنامج باستخدام لغة C# بالإصدار 3.0 حيث هذه اللغة توفر إمكانيات كبيرة في التفاعل مع النظام وملفاته وكذلك توفير الأدوات اللغوية لتصميم واجهات تفاعلية لتسهيل عملية الاتصال بين الحاسوب والمستخدمين وكذلك لعرض النتائج للمستخدمين بكافة مستوياتهم.

يتكون البرنامج من جزئين أساسيين:

الجزء الأول: برنامج الINDEX، هذا البرنامج يعمل على جلب وتخزين ملفات الحاسوب في ملف فهرسة خاص بمحرك البحث SPCI، والتي سوف تستخدم لاحقاً في عملية البحث وجلب النتائج المطلوبة، وهذا البرنامج هو

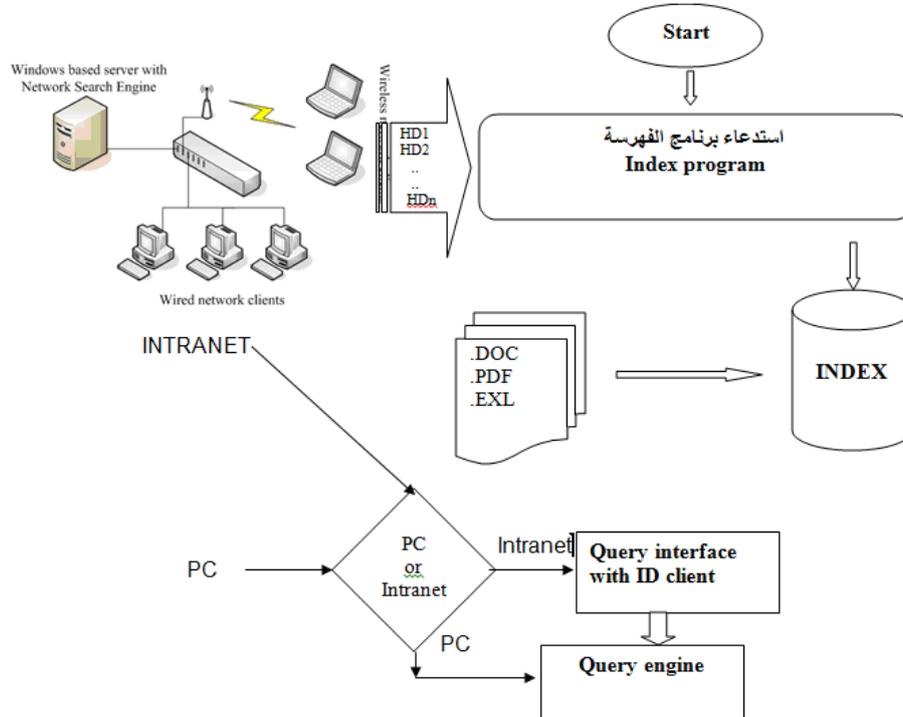
عبارة عن برنامج فرعي indexer قمنا ببنائه ضمن أجزاء محرك البحث، حيث يقوم بعملية بناء ملف فهرسة في أول تشغيل للبرنامج وكذلك عند حصول تغيير في الملفات الموجودة في الحاسوب يقوم بعملية إعادة تكوين ذلك الملف، وبالطريقة نفسها على بقية الحواسيب المتواجدة في الشبكة كما في شكل (2).

الجزء الثاني: يتعلق هذا الجزء بعملية البحث، وهو الواجهة التي يتحاور معها المستخدم ويطلب منها ما يحتاج إليه من ملفات في صندوق الحوار الخاص بعملية البحث، إذ قمنا بتجزئة العملية إلى ثلاثة أجزاء أساسية:

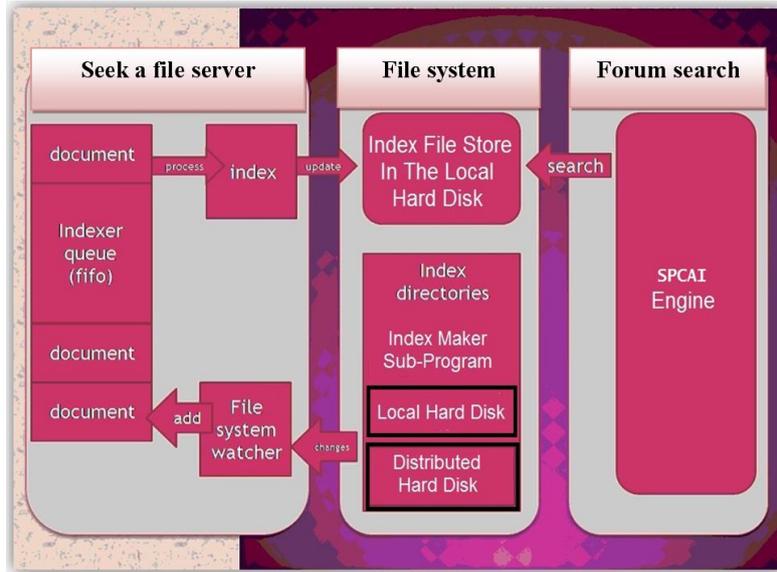
- 1- عملية الـ query parser.
- 2- عملية الـ query engine.
- 3- عملية عرض نتائج البحث المطلوبة.

1-7 تكوين الـ INDEX

إن عملية تكوين الـ Index مهمة جداً، فمن دونها لا يستطيع البرنامج الوصول إلى أي ملف من الملفات المخزونة، وسوف تكون عملية البحث صعبة جداً، ففي حالة تشغيل برنامج البحث لأول مرة يجب علينا أن نقوم بعملية بناء الـ index للبرنامج من خلال استدعاء برنامج فرعي الـ indexer، حيث يقوم هذا البرنامج ب جلب الملفات الصورية بكافة أنواعها (.jpg, gif, bmp,)، وكذلك الملفات النصية ذات الامتدادات (pdf, htm,) (ppt, doc, txt, xls)، من القرص الصلب (أو الأقراص في حالة الشبكة)، وملف فهرسة خاص بالمحرك (المحركات) يخزن داخل امتداد معين قمنا بتحديدده في مسار خاص لحمايته من أي تطفل من قبل المستخدم. الشكل (3) يوضح عمل البرنامج الفرعي بشكل تفصيلي، وبعد قيامنا بعملية بناء index للبرامج نستطيع الآن إجراء عملية البحث بالاعتماد على ملف الفهرسة الذي قمنا ببنائه من خلال البرنامج الفرعي الـ indexer.



شكل (2). بناء Index لشبكة انترانيت.



الشكل (3). عملية تكوين ملف الفهرسة

2-7 عملية البحث Searching

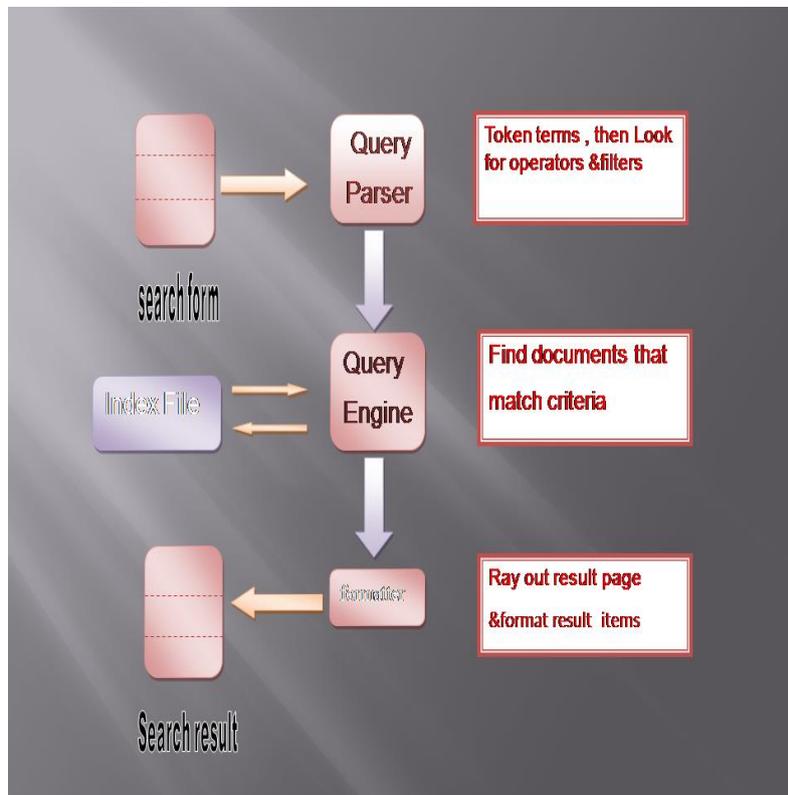
تقسم عملية البحث إلى ثلاثة أجزاء:

أ- عملية query parser.

ب- عملية query engine.

ج- عملية عرض ناتج البحث المطلوب.

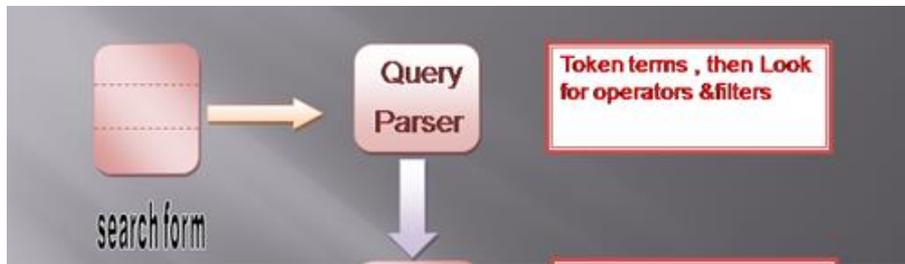
ويمكن توضيح هذه العمليات في الشكل (4)



الشكل (4). خطوات عملية البحث

7-2-1 عملية الـ Query Parser

تتم عملية تحديد الاستعلام من خلال برنامج فرعي قمنا بإنشائه اعتماداً على قاعدة بيانات صغيرة خاصة بالـ Keyword & Identifier، إذ أنه عند استدعاء هذا البرنامج الفرعي، في البداية يقوم بتقطيع النص إلى مفاتيح تسمى الـ Tokens، وبعد عملية تقسيم النص إلى الـ Tokens، يقوم بعملية مقارنة كل Token مع قاعدة البيانات الموجودة في البرنامج، إذا كانت هذه الـ Token مطابقة مع البيانات الموجودة في قاعدة البيانات، في هذه الحالة يقوم query parser بتحليل هذه الـ Token ومعرفة صيغة السؤال والنتيجة المطلوبة منه، أي هل المستخدم يبحث عن ملفات نصية أم صوتية، ومن أي نوع بالتحديد. أما في حالة عدم حصول مطابقة ما بين الـ Token مع بيانات التابعة لقاعدة البيانات في هذه الحالة يقوم الـ query parser ببناء قاعدة بيانات مصغرة تدعى بـ token_search حيث تضم الـ Token المراد البحث عنها، عندما ينتهي query parser من عملية اختبار الـ Token، يتم إمرار قاعدة البيانات token_search التي تضم الـ Token المراد البحث عنها إلى البرنامج الفرعي indexer، ويمكن توضيح أعلاه في الشكل (5).



الشكل (5). عملية تحديد الاستعلام

7-2-2 عملية الـ Query Engine

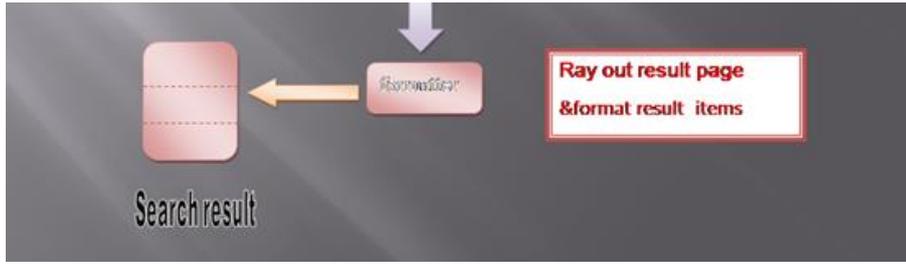
يقوم هذا البرنامج بعملية البحث عن كل Token الموجودة في قاعدة البيانات بالاعتماد على قاعدة البيانات التي تم تكوينها باستخدام البرنامج الفرعي الـ indexer، وفي حالة إيجاد البيانات المطلوبة يقوم هذا البرنامج ببناء قاعدة بيانات تضم البيانات المطلوب البحث عنها (أي يخزن المسار الكامل للبيانات المطلوبة في قاعدة البيانات). ويمكن توضيح أعلاه في الشكل (6).



الشكل (6). تحليل الاستعلام

7-2-3 عملية عرض النتائج

في هذه العملية يتم عرض البيانات المطلوبة للبحث وكذلك عرضها لشاشة العرض، كما موضح بالشكل (7).



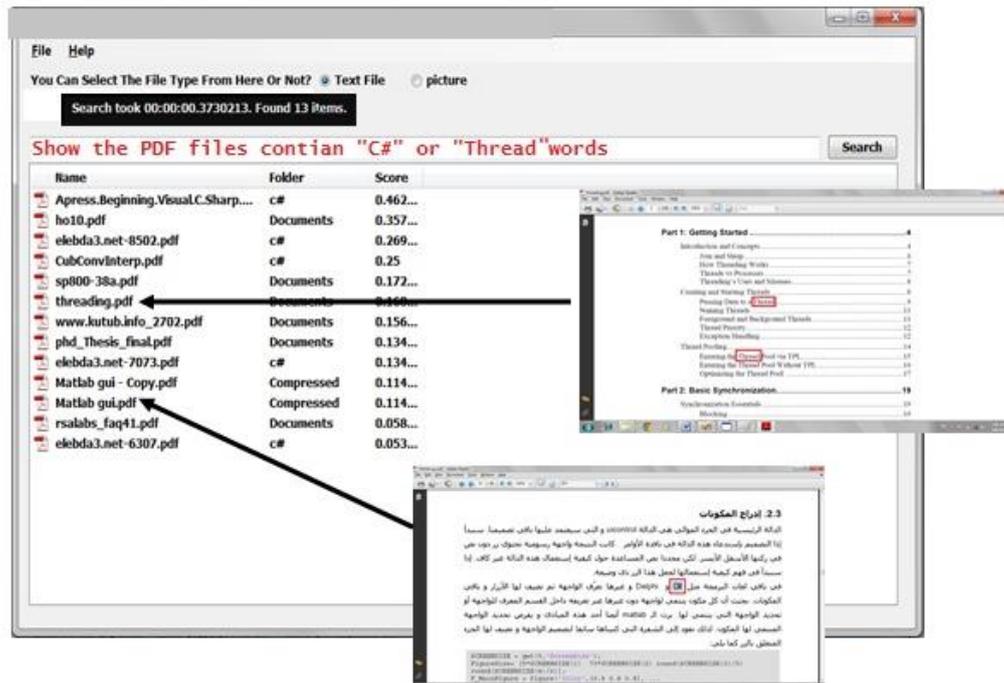
الشكل (7). عرض النتائج

وفيما يلي مثال تطبيقي:

البحث عن ملفات من نوع PDF تحتوي على كلمة "C#" أو كلمة "Thread"، نلاحظ سؤال الاستعلام سيذكر في صندوق الاستعلام كما موضح بالشكل (8):

Show the PDF file contain "c#" or "Thread" words

تمت مقارنة الوقت للمثال أعلاه بطريقة البحث التقليدي مع البحث باستخدام البرنامج ولوحظ أن الوقت المستغرق للحصول على الملفات عند استخدام برنامج (SPCI) هو 00:00:00.3730213 في حين أنه في طريقة البحث التقليدية هو 00:00:3.530213. وتمت عملية حساب عامل الوقت بإضافة دالة الوقت في البرنامج لتظهر لنا الوقت كما في الشكل (8). في حين أنه في البحث التقليدي الاعتماد على وقت الحاسبة في الحصول على النتائج.



الشكل (8). مثال تطبيقي

8- النتائج

تم تطبيق برنامج (SPCI) وتنفيذ جميع الخيارات وكانت النتائج كالآتي:

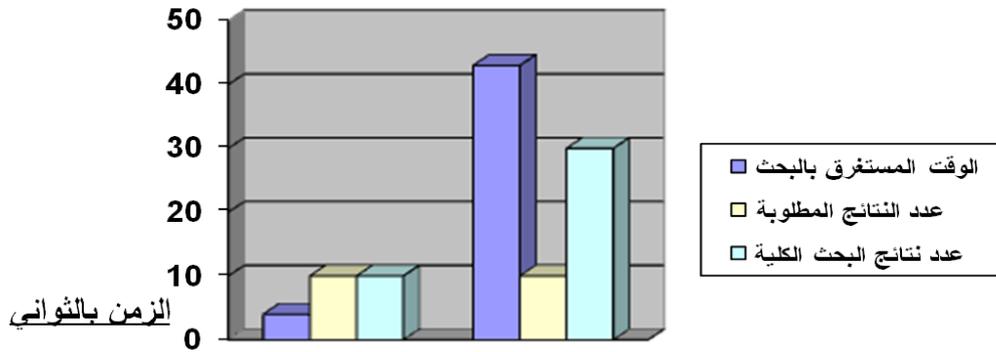
1- يوفر محرك البحث SPCI السرعة العالية في استرجاع الملفات نوع Doc ليس فقط بالاعتماد على اسم ومسار الملف وإنما بمحتوى هذه الملفات عند استخدام ذلك المحتوى بوصفها مفاتيح للبحث عن الملف المطلوب.

2- يوفر محرك SPCI دقة وكفاءة عالية في الحصول على البيانات المطلوبة من خلال تقليص عدد الملفات المسترجعة والسبب يعود في ذلك بالاعتماد على معلومة محددة كمفتاح الاستعلام على العكس من أنظمة البحث الدلالية (Directories) التي تقوم باسترجاع جميع الملفات عن طريق الاسم أو المسار، ويمكن ملاحظة هذا الأمر من خلال الشكل (9) الذي يوضح الفرق بين عملية البحث في محرك البحث SPCI ومحرك البحث الخاص بنظام التشغيل حول ملفات من نوع PDF.

3- إضافة خاصية جديدة لنظم التشغيل وهي التناوب مع المستخدم باللغة الطبيعية وذلك عند الاستعلام عن ملف يجهل المستخدم اسمه أو مساره.

4- يوفر برنامج SPCI إمكانية البحث عن ملفات نوع (TXT, PDF, Word, Excel, Html) بالاعتماد على فهرسة جديد، على العكس من برنامج البحث التقليدي الموجود في أنظمة التشغيل الويندوز التي تعتمد على الفهرسة الهرمية التي يوفرها نظام التشغيل.

5- يمكن فهرسة المجلدات المتواجدة على الشبكة في جميع محركات الأقراص أو اختيار أحدها.



الشكل (9). مقارنة النتائج بين البحث في SPCI مع البحث التقليدي في Windows

9- الاستنتاجات

محركات البحث هي عبارة عن برامج مساعدة لتسريع استرجاع المعلومات المخزونة على كل نظم الحاسوب وهذه النظم أما أن تكون شبكة عالمية Word Wide Web ويسمى هذا البحث عن المعلومات في الشبكة العامة وهو النوع الأول، أو نظم شبكات داخلية (Intranet) وتسمى محركات البحث في المؤسسات وهو النوع الثاني والأخيرة على الحواسيب الشخصية المنفردة. حقق برنامج (SPCI) إمكانية البحث من النوع الثاني والثالث على بيانات نوع (TXT, PDF, Word, Excel, Html) وتحقق معيارين أساسيين هما السرعة والدقة كما موضح في الشكل (8). لقد تم اعتماد اللغة الطبيعية في التناوب مع الحاسوب من خلال صندوق الحوار في التعرف على البيانات المطلوبة إذ اعتمد البرنامج على استخدام كلمات أساسية لاسترجاع الملفات (KEYWORD).

المصادر

- [1] محمد عبد الهادي، زين الدين، "محركات البحث على الانترنت للمكتبات ومراكز المعلومات"، 2007، ايبيس.كوم للنشر والتوزيع، ص3-5، 8-9، 12-14.
- [2] ربيع سيد إبراهيم، "محركات بحث عن الصور الثابتة على الانترنت"، دراسة تحليلية، 2007، مكتبة الملك فهد الوطنية، السلسلة الثانية (51)، ص 70، 71، 72.
- [3] محجوب حمد عبد الرحمن، "محركات البحث"، ص3، 4.
- [4] "The Structure of Search Engine Law", 2006, New York Law School Inc., No. 05 Civ. 6487, U.S. Dist. LEXIS 3114;
- [5] "Search Engine Content Analysis", 2008, School of Software Engineering and Data Communications, Faculty of Information Technology, Queensland University of Technology, Brisbane, Australia.
- [6] www.r-7oob.com/vb/archive/index.php/t-110872.html
- [7] <http://www.swalif.net/softs/swalif12/softs209192>
- [8] <http://www.tech-faq.com/depth-first-search-algorithm.html>
- [9] Huaiyu Zhu, Sriram Raghavan, Shivakumar Vaithyanathan and Alexander Loser, "Navigating the Intranet with High Precision", 2007.
- [10] "University search engine", 2010, Journal of Kufa for mathematics and computer, Vol.1, No. 2, Oct, 2010, PP 46-53.