كفاءة محركات البحث العربية في استرجاع المعلومات

دراسة تقويمية مقارنة لآليات البحث

عبدالله بن إبراهيم المبرز ❖

المقدمة :

منذ أتيحت شبكة الإنترنت للاستخدام العام والمعلومات المنشورة عليها تتزايد بشكل مستمر ومتسارع، حتى أنه بات من المستحيل تتبع كل ما يستجد من إضافات على الإنترنت نظراً لتضاعفها بشكل لم يكن ممكناً من قبل. حيث قدر فينج وبهراجاف وبينوك Feng, Bhargava, & Pennock (٢٠٠٧) أن ما كشف من الصفحات عام ١٩٩٩م نحو ٨٠٠ مليون صفحة، تحتوي على ستة تيرابايت من النصوص ومخزنة على ٢,٨ مليون خادم. ثم تضاعفت مصادر المعلومات المفتوحة على الإنترنت حتى وصلت في الوقت الحالي إلى بلايين الصفحات المكشفة مخزنة على ٤٠ مليون خادم. هذه الضخامة من المعلومات والانفجار المعرفي على الشبكة كان له أثر سلبي في التعرف إلى الوثائق المناسبة. نظراً لأنه لم يعد بإمكان أي شخص أو باحث الحصول على ما يحتاجه من محتويات الإنترنت رغم كثرة المعلومات المتاحة وتنوع مصادرها دون استخدام أدوات مساندة للوصول إلى المواقع التي تناسب احتياجاته المعلوماتية.

لذا فقد ظهرت أدوات مختلفة للتنظيم والاسترجاع على الإنترنت تمكن المستخدم من تأمين بغيته من المعلومات. وقد ظهرت تلك الأدوات في أشكال مختلفة، منها محركات البحث، والأدلة الموضوعية، والبوابات. ولا ريب أن محركات البحث العامة والموضوعية ليست

❖ بكالوريوس مكتبات ومعلومات من قسم المكتبات والمعلومات، كلية الآداب، جامعة الملك سعود عام١٤١٠هـ.
– ماجستير في المكتبات والمعلومات، جامعة الإمام محمد بن سعود الإسلامية عام ١٤١٧هـ.
– دكتوراة في علم المعلومات، جامعة شمال تكساس – الولايات المتحدة الأمريكية عام ١٤٢٨هـ.
– يعمل حاليًا رئيسًا لقسم المعلومات بكلية الحاسب الآلي والمعلومات، بجامعة الإمام محمد بن سعود الإسلامية.

أبرز تلك الأدوات استخداماً فحسب، بل إنها الأكثر شيوعًا بين تطبيقات الإنترنت كافة. حيث ذكر ريني وشرماك Rainie & Shermak (٢٠٠٥م) أنها تأتي في المركز الثاني بعد استخدام البريد الإلكتروني. كما توصل فينج وبهراجاف وبينوك Feng, Bhargava, & Pennock (٢٠٠٧) إلى أن محرك البحث الشهير جوجل Google يستقبل يومياً أكثر من ٢٥٠ مليون استفساراً.

ورغم أن المحركات لا تعتمد على الجهد البشري في التحليل الموضوعي بل تستخدم برمجيات بحث خاصة، إلا أنها اكتسبت هذه الأهمية كونها تغطي مساحة شاسعة من محتويات الإنترنت وتعتمد على خوارزميات متقدمة لتكشيف المواقع التي تقوم بزيارتها. كما أنها تتنوع من ناحية تخصصها الموضوعي والشكلي في الاسترجاع، كي تلبي اهتمامات شرائح متباينة من المستفيدين.

ولعل من أكبر التحديات التي تواجه محركات البحث هو كيفية إرضاء المستخدم وتلبية احتياجاته المعبر عنها من خلال كلمات محدودة أو جمل قصيرة قد تخفي وراءها معاني خاصة، لم يستطع المستفسر صياغتها بطريقة يفهمها النظام. ورغم تلك التحديات إلا أن محركات البحث ظلت وما زالت من الأدوات الأساسية التي يعتمد عليها في البحث عن المعلومات على الشبكة العالمية. ويلحظ حالياً أن المتخصصين في استرجاع المعلومات والشركات المهتمة بهذا الشأن في بحث مستمر لإيجاد محركات مثالية، أو تطوير القائم منها كي تستطيع السيطرة على المحتوى الرقمي التراكمي للإنترنت. ومع التطور والتقدم في أداء محركات البحث إلا أنه لم يوجد بعد محرك بحث دقيق وشامل يستطيع تكشيف محتويات الإنترنت كافة لعوامل اقتصادية وفنية مختلفة. فمن الناحية الاقتصادية، فإن من المكلف مادياً إجراء مسح شامل لمحتويات الإنترنت بما يستجد فيها من صفحات بشكل يومي. لأن تحقيق ذلك يتطلب شبكة ضخمة من الخوادم التي تحتوي على مئات أو حتى آلاف من أجهزة الحاسب الآلي المخصصة لهذا الغرض. أما التحديات الفنية فتظهر من صعوبة التعرف إلى كل الصفحات المرتبطة باستفسارات البحث في بيئة رقمية ضخمة لا تعرف حدودها وحجمها؛ مما جعل من الصعب حالياً قياس حجم التغطية لمحركات البحث القائمة **(ليواندوسكي Lewandowski، ٢٠٠٥م)**.

مشكلة الدراسة :

رغم قلة المحتوى العربي على الإنترنت، إلا أنه كغيره من الإنتاج الفكري في اللغات

الأخرى قد تضاعف بدرجات مرضية إلى حد ما، بسبب مرونة النشر الإلكتروني وبعده عن الإجراءات الروتينية التي يتصف بها عادة النشر التقليدي العربي. وقد ظل الباحث العربي لفترة طويلة ولا يزال يعتمد على المحركات الأجنبية للوصول إلى مبتغاه من المعلومات المنشورة على المواقع العربية. وقد زاد ذلك الاستخدام مع ظهور المحركات التي تدعم اللغة العربية فكانت بديلاً غير كافٍ عن ضرورة وجود محركات بحث عربية مصممة خصيصاً لمسح وتكشيف المواقع العربية. وذلك لطبيعة اللغة العربية واختلاف بعض خصائصها عن خصائص اللغات الأخرى، حيث إن اللغة العربية تختلف عن اللغات الأخرى في أن لها خصائص دلالية وتركيبية معقدة تؤثر على فعالية الاسترجاع ودقته .

وقد ظهرت بناء على تلك الحاجة محركات بحث عربية في محاولة لسد تلك الفجوة اللغوية، التي تعتمد على تحليل البناء اللغوي لمحتوى المواقع كي تكون قادرة على فهم الاستفسارات الموجهة باللغة العربية. ورغم ظهور تلك المحركات إلا أن ثمة انطباعاً سائداً بأنها غير قادرة على منافسة المحركات الأجنبية التي تدعم البحث باللغة العربية، فكان ذلك سبباً في عدم شيوع استخدامها، ومواجهتها لكثير من المشكلات التي لم تحقق رضا الباحث العربي.

ورغم أن تصميم محرك بحث في حد ذاته ليس عملية معقدة، إلا أن تصميمه بشكل خاص لاسترجاع النصوص العربية يعد من الأمور الشائكة التي عانى منها الباحثون العرب منذ التفكير بتصميم نظم استرجاع خاصة باللغة العربية، وذلك لطبيعتها، وتشعب الصعوبات المرتبطة بها، فثمة صعوبات عامة مثل المرونة وتعدد المفردات وتفاوتها، وهناك صعوبات تتعلق بإشكالات صرفية أو نحوية وأهمية التقيد بالشفرة العربية الموحدة في تمييز أشكال الألف ومواضع الهمزة وغيرها من الإشكالات **(الفرغولي، ٢٠٠٢م)**.

والمتتبع لمحركات البحث العربية والدراسات المرتبطة بها يرى أنها تعاني من مشكلات متعددة لم تسمح لها بالانتشار. ويمكن تلخيص بعض تلك المشكلات فيما يلي:

١- عدم الاستمرارية والتوقف بعد فترة قصيرة من تطبيقها حيث ظهرت محاولات كثيرة لم يكتب لها النجاح، وقليلة هي المحركات التي استمرت.

٢- عدم وجود الدعم المناسب لها سواء فنياً أو مالياً أو بشرياً.

٣- اعتماد الكثير منها على الجهود الفردية ومبادرات الأشخاص، حيث لا تتوافر لها

الرعاية من قبل المؤسسات المهنية والشركات المتخصصة.

٤- بعضها لا يعتمد على آليات بحث خاصة به، وإنما يعمل كواجهة فقط تخفي وراءها محركات بحث أجنبية.

وبناء على ما تعانيه محركات البحث العربية من عدم القدرة على كسب ثقة المستخدم العربي، فقد رأى الباحث ضرورة تقويم أداء بعض المحركات الموجودة لقياس مدى قدرتها على التعامل مع الاستفسارات المطروحة باللغة العربية، لأجل الخروج بتوصيات يمكن أن تسهم في معالجة ضعف أدائها وتحقيق الهدف من وجودها.

أسئلة الدراسة:

لتحقيق الأهداف من وراء إجرائها، فإن هذه الدراسة تحاول الإجابة عن التساؤلات التالية:

١- ما مدى الارتباط بين المواقع المسترجعة والاستفسارات الموجهة لمحركات البحث العربية؟

٢- ما أفضل الأساليب البحثية التي تستطيع محركات البحث العربية التعامل معها من بين الأشكال التالية: اللغة الطبيعية، الربط البولياني، الكلمات المفتاحية؟

٣- ما مدى قدرة محركات البحث العربية في الاسترجاع مقارنة بمحركات البحث الأجنبية؟

٤- ما المشكلات التي تعاني منها محركات البحث العربية؟

الدراسات السابقة:

نظرًا لأهمية محركات البحث كأداة للوصول إلى المعلومات والمواقع المختلفة على الإنترنت، ولكثرة استخدامها من قبل شريحة كبيرة من المستخدمين، فقد اجريت دراسات متعددة لتقويمها ومقارنتها بعضها مع بعض؛ لأجل التعرف إلى أبرز تلك المحركات، إضافة إلى معرفة جوانب القوة ومواطن الضعف فيها.

فقد قام جوتشيم جريسبام Joachim Griesbaum (٢٠٠٤م) بدراسة مقارنة لثلاثة محركات بحث تدعم اللغة الألمانية وهي ألتافيستا AltaVista وجوجل Google وليكوس Lycos. وكان الهدف من الدراسة قياس مدى دقة الاسترجاع باللغة الألمانية من خلال تحليل العشرين نتيجة الأولى لعشرين استفساراً موجهاً لتلك المحركات. وقد توصل الباحث إلى أن محرك البحث جوجل Google حصل على دقة استرجاع أفضل من المحركين الآخرين، يليه ليكوس Lycos، وأخيراً ألتافيستا AltaVista.

وفي دراسة أخرى قام باريلان ولفين وماتحسان Bar-Ilan, Levene & Mat-Hassan (٢٠٠٦م) بتتبع التغيرات التي تحدث في ترتيب العشر النتائج الأولى لمحركي البحث AlltheWeb وجوجل Google من خلال طرح مجموعة من الاستفسارات في فترتين زمنيتين مختلفتين. وقد لاحظ الباحث أن ترتيب النتائج في محرك AlltheWeb كان ثابتاً خلال التجربتين رغم اختلاف الفترة الزمنية. في حين ظهرت تغيرات في ترتيب النتائج لمحرك جوجل Google. كما لاحظ باريلان انخفاضًا كبيرًا في معدل تكرار النتائج بين المحركين؛ مما يعطي مؤشراً أن آلية البحث والتكشيف بينهما مختلفة.

كما أجرى ليون فايون Liwen Vaughan (٢٠٠٤م) دراسة مقارنة لترتيب نتائج البحث التي تقوم بها آلياً ثلاثة محركات بحث وهي جوجل Google، وألتافيستا AltaVista، وتيوما Teoma، مع ترتيب آخر قام به مستخدمين فعليين، من أجل التعرف إلى مدى دقة محركات البحث في تحديد أولويات الارتباط الموضوعي بين الاستفسار المطروح والنتائج المسترجعة. وقد توصل الباحث إلى أن جوجل حصل على معدلات عالية في نسبة التوافق بين ترتيب المستخدمين وترتيب محرك البحث لديه. في حين حصل المحركان الآخران على نسبة توافق متدنية.

كما درس أليريزا إسفانديار موغادام Alireza Isfandyari Moghaddam (٢٠٠٧م) محركات البحث المتعدد Metsearch engines وهي التي تقوم بالبحث في محركات بحث مختلفة عند طرح استفسار واحد عليها. وتمت الدراسة على عشرين محركاً من ذلك النوع، توصل فيها الباحث إلى أن تلك المحركات تقدم دقة استرجاع عالية، ويمكن أن تكون بديلاً جيداً لمحركات البحث العامة. وبناء عليه فقد أوصى الباحث بأن على المستخدم أن يكون واعياً بسلبيات الاعتماد على محرك بحث واحد، وأن ذلك قد يؤدي إلى فقدان مواقع ومعلومات مهمة قد تكون أعطيت تصنيفاً عالياً في الترتيب من قبل محركات بحث أخرى.

كما أن مقارنة محركات البحث على الإنترنت مع قواعد المعلومات المتخصصة لم تغفل من الدراسة من قبل المتخصصين في المجال. حيث قامت ماري شلتز Mary Shultz (٢٠٠٧م) بمقارنة محرك البحث جوجل العلمي Google Scholar مع قاعدة البيانات PubMed المتخصصة في العلوم الطبية. وقد اختار الباحث جوجل العلمي لتغطيته لوثائق علمية محكمة

تشتمل على المجلات العلمية، الرسائل الجامعية، الكتب، الملخصات. وعلى الرغم من أن واجهة البحث الخاصة بمحرك البحث جوجل أسهل استخداماً، إلا أن الباحث وجد أن واجهة البحث لقاعدة PubMed تقدم اختيارات وخصائص أكثر وأدق للمستخدم. كما توصل إلى أن PubMed بخلاف جوجل تطبق استراتيجية البحث بالكلمات المقيدة لاعطاء الباحث المجال لتركيز موضوع بحثه أو تضييقه.

ومن الدراسات العربية في هذا الجانب ماقامت به نادية مصطفى العيدروس أحمد (٢٠٠٦م) من قياس مدى كفاءة محركات البحث العربية والعالمية الداعمة للغة العربية في استرجاع المعلومات العربية. وقد اعتمدت الباحثة على محركي بحث عالميين هما جوجل Google وألتافيستا AltaVista وآخرين باللغة العربية هما أين والإدريسي. وقد خرجت الباحثة بمجموعة من النتائج أبرزها أن محركي جوجل Google وألتافيستا AltaVista قد حصلا على أعلى درجة من الاستفادة الفعلية من النتائج المسترجعة، بينما قلت هذه النسبة في محركي أين والإدريسي. كما وجدت الباحثة أن عدد الوثائق المكررة في قائمة نتائج البحث الواحد تنعدم في محركات البحث العربية، بينما ترتفع في محركات البحث العالمية.

كما حاول خالد عبدالفتاح محمد (٢٠٠٧م) قياس كفاءة ستة محركات بحث في تكشيف واسترجاع المعلومات والمواقع العربية على الإنترنت. وقد توصل إلى أنه كلما زاد عدد كلمات الاستفسار، ارتفعت معدلات التداخل والتكرار بين محركات البحث. وكلما تكرر تكشيف الوثيقة من جانب أكثر من محرك بحث واحد، ارتفعت معدلات صلاحية الوثيقة، وارتفع معها ترتيب الوثيقة ضمن مجموعة الوثائق المسترجعة.

منهجية الدراسة :

ثمة معايير مختلفة لتقويم محركات البحث تختلف من دراسة إلى أخرى حسب أهداف ذلك التقويم. فمن الزوايا التي يمكن النظر إليها في هذا الجانب سعة التغطية، وحداثة الصفحات المكشفة، وسهولة الاستخدام، والارتباط الموضوعي باستفسارات البحث. ورغم أهمية تلك الجوانب إلا أن الباحث ركز في تقويم محركات البحث محل الدراسة على دقة نتائج الاسترجاع وتحديد مدى ارتباطها بالاستفسارات الموجهة من قبل المستخدمين.

وقد اعتمد الباحث على ثلاثة مناهج علمية لتحقيق أهداف الدراسة والإجابة عن أسئلتها وهي:

١- المنهج التجريبي: وذلك عن بتجربة مدى جودة أداء محركات البحث المدروسة عن طريق توجيه مجموعة من الاستفسارات بصيغ مختلفة.

٢- منهج تحليل المحتوى: وتم عن طريق الاطلاع المباشر على النتائج العشر الأولى لكل استفسار تم طرحه، وتحليل محتواها الموضوعي لتحديد مدى ارتباطها بالاستفسارات المطروحة. وقد بلغ مجموع المواقع التي تم فحصها ٥٤٠ موقعاً استرجعت من خلال ١٨ استفساراً تم طرحها على محركات البحث الثلاثة.

٣- المنهج المقارن: وذلك بمقارنة نتائج محركي البحث العربيين قيد الدراسة مع محرك أجنبي يدعم البحث باللغة العربية لمعرفة الفروقات بين هذه النتائج والخروج بتصور عن موقع محركات البحث العربية بين المحركات الأجنبية.

إجراءات التحليل:

لتحقيق التقويم المطلوب لمحركات البحث ومقارنة أدائها معًا فقد تم صياغة ١٨ استفساراً تنتمي إلى موضوعات مختلفة تم تقسيمها إلى ثلاث مجموعات كل واحدة منها تحتوي على ستة استفسارات تشترك فيما بينها في اسلوب صياغة موحد. وفيما يلي شرح لهذه المجموعات:

١- المجموعة الأولى: ويندرج تحتها الاستفسارات التي وجهت لمحركات البحث بين علامتي تنصيص او مايطلق عليه مطابقة العبارة أو الجملة Phrase ، وذلك لاسترجاع المواقع التي تحتوي على الجملة نفسها الموجودة بين العلامتين.

٢- المجموعة الثانية: وتحتوي هذه المجموعة على الاستفسارات المطروحة باللغة الطبيعية Natural Language، وهي التي لا تحتوي على علامات تنصيص أو روابط بولیانية. كما أنها تحتوي على أكثر من كلمة في الاستفسار الواحد.

٣- المجموعة الثالثة: وهي الاستفسارات التي صيغت باستخدام الروابط البولیانية Boolean Operation وهي: أو، و، ليس. وذلك لتضييق موضوع البحث أو توسيعه؛ للحصول على نتائج لها ارتباط دقيق بالموضوع المطلوب. وتقوم مهمة الأمر (و) على ربط كلمتين أو أكثر في الاستفسار بحيث تسترجع فقط المواقع التي تحتوي على هذه الكلمات مجتمعة. في حين أن مهمة الأمر (أو) توجيه محرك البحث لاسترجاع المواقع التي تحتوي على أي من الكلمات الموجودة بين الرابط. وتعمل (ليس) على استبعاد المواقع التي تحتوي على الكلمات التي ترد بعدها.

ويوضح الجدول رقم (١) جميع الاستفسارات التي قدمت لمحركات البحث لأجل التقويم، مع بيان المجموعة التي تنتمي إليها.

الجدول رقم (١)

الاستفسارات المطروحة على محركات البحث

رقم الاستفسار	رقم الاستفسار	تصنيف المجموعة
١	"خدمات المعلومات الرقمية"	المجموعة الأولى
٢	"التعليم العالي في المملكة العربية السعودية"	
٣	"سيرة الخليفة عمر بن الخطاب"	
٤	"تاريخ الدولة السعودية الأولى"	
٥	"ظاهرة العنوسة"	
٦	"العلاج بالأعشاب"	
٧	الثروة السمكية في عمان	المجموعة الثانية
٨	انتشار المخدرات بين الشباب	
٩	المجاعة في أفريقيا	
١٠	استخدام الإنترنت في الوطن العربي	
١١	الوقاية من مرض الإيدز	
١٢	حقوق المرأة العربية	
١٣	صناعة و "الدخان أو السجائر"	المجموعة الثالثة
١٤	أركان و الإسلام	
١٥	"المهاجرون العرب" و الولايات المتحدة ليس أوروبا	
١٦	التبادل و "التجاري أو الاقتصادي" و العربي	
١٧	"صناعة أو إنتاج" و "النفط أو البترول"	
١٨	سيرة والخلفاء والراشدين	

وبعد طرح تلك الاستفسارات على محركات البحث قيد الدراسة، تم تقويم العشر النتائج الأولى لكل استفسار على اعتبار أن المستفيد غالباً ما ينظر إلى العشر النتائج الأولى من قائمة النتائج المسترجعة (**جيكبس جاريد ومات روبنز Jared Jacobs & Matt Rubens، ٢٠٠٣م**). وقد تم التقويم بناء على مقياس حدد مسبقاً لتوضيح مدى ارتباط النتيجة بالاستفسار المطروح. وقد تم صياغة المقياس على أساس إعطاء رموز حرفية للنتائج المسترجعة، وكل حرف يعبر عنه بدرجات توضح مدى الارتباط بين الاستفسار والنتيجة. وفيما يلي شرح تفصيلي لهذه الرموز مع بيان مقاييسها ومعايير التقييم لكل نتيجة:

الرمز (أ): وقد أعطيت النتيجة التي حصلت على هذا الرمز ٥ درجات، وهي الدرجة العليا في المقياس، وتعنى الارتباط الفعلي بين استفسار البحث والنتيجة المسترجعة. ونظراً لاختلاف طبيعة كل مجموعة عن الأخرى، فتحديد ما يقصد بالارتباط الفعلي يختلف بين المجموعات. ففي حالة المجموعة الأولى تعطى هذه الدرجة عندما تظهر العبارة المستفسر عنها بين علامتي التنصيص ضمن ثنايا الموقع المسترجع. أما في المجموعة الثانية فإن هذه الدرجة تخصص للمواقع التي تحتوي على معلومات تفصيلية مرتبطة موضوعياً باستفسار البحث. وفي المجموعة الثالثة تعطى هذه الدرجة للنتائج التي تتطابق مع استفسار البحث المصاغ باستخدام الروابط البولينية، ولها في الوقت نفسه علاقة موضوعية بذلك الاستفسار.

الرمز (ب): وقد أعطيت النتائج التي حصلت على هذا الرمز ٣ درجات، وتعني أن النتائج لها ارتباط جزئي وتحتوي على معلومات غير تفصيلية عن استفسار البحث. كما تندرج تحت هذه الدرجة النتائج التي تحتوي على روابط لصفحات تنمي للفئة الأولى، أو معلومات ببليوجرافية عن مصادر تغطي موضوع الاستفسار المطروح.

الرمز (ج): وقد أعطيت النتائج التي حصلت على هذا الرمز درجة واحدة، وتعطى للنتائج التي ليس لها علاقة باستفسار البحث، ولكنها استرجعت نظراً لورود بعض كلمات الاستفسار أو كلها في الموقع، دون أن يكون هناك ارتباط موضوعي مع الاستفسار.

الرمز (د): وقد أعطيت النتائج التي حصلت على هذا الرمز صفر. وتعطى في حالة استرجاع روابط معطلة أو مواقع مغلقة.

إضافة لما سبق فهناك درجة أعطيت للمواقع المكررة، وتم حسابها عن طريق تقسيم درجة الموقع الذي ظهر اولاً بعدد مرات تكراره بحيث تكون المعادلة كما يلي:

(درجة الموقع الأول ÷ عدد مرات التكرار).

عينة الدراسة:

أجريت الدراسة على محركي بحث عربيين منتقيين وهما محرك البحث (أين)، ومحرك البحث (عربي)، ومحرك بحث أجنبي واحد وهو ألتافيستا AltaVista. وقد تم اختيار المحركين العربيين على أساس أنهما الأشهر من بين غيرها من المحركات العربية. ولأنهما صمما من قبل مؤسسات متخصصة بعكس الكثير من المحركات العربية على الإنترنت والتي صممت من قبل أفراد. كما تم اختيار محرك

البحث ألتافيستا AltaVista لأنه من المحركات الأجنبية المتميزة التي تدعم البحث باللغة العربية.

مناقشة النتائج:

يتناول هذا الجزء من الدراسة تحليل البيانات التي تم جمعها من أجل الوصول إلى إجابات واضحة لأسئلة البحث، والتعرف إلى مدى كفاءة الاسترجاع في محركات البحث المدروسة، والمقارنة بينها للتعرف إلى أكثرها قدرة ودقة في إمداد الباحثين بالمعلومات المرتبطة باحتياجاتهم المعلوماتية.

محرك البحث (أين):

وتم في البدء تحليل النتائج المسترجعة من محرك البحث (أين) باستخدام الاستفسارات المقسمة إلى ثلاث مجموعات وهي البحث بمطابقة العبارة، والبحث باللغة الطبيعية، والبحث بالربط البولياني. ويوضح الجدول رقم (١) النتيجة النهائية لذلك التحليل.

الجدول رقم (١)

درجة التقويم لمحرك البحث (أين) حسب نوع استراتيجية البحث

نوع الاستفسار	الدرجة
مطابقة العبارة	٢,٣١
اللغة الطبيعية	٢,٤١
الربط البولياني	٢,٥٠

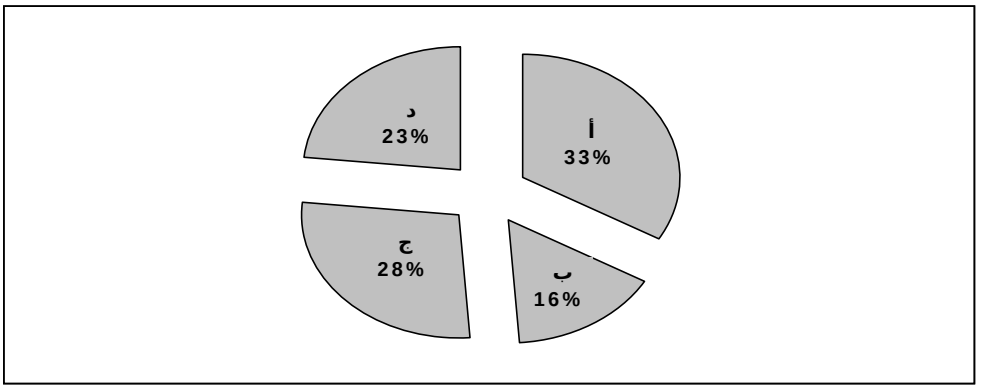
ويلحظ من خلال قراءة الجدول السابق أن قدرة محرك (أين) في استرجاع المواقع ذات الصلة باستفسار البحث منخفضة في الأنواع الثلاثة من الاستفسارات. حيث حصل الاستفسار باستخدام الربط البولياني على ٢,٥٠ درجة من ٥ كأعلى درجة بين المجموعات الأخرى. يليه البحث باللغة الطبيعية وحصل على ٢,٤١ درجة، ثم البحث بمطابقة العبارة وحصل على ٢,٣١ درجة. وتشير هذه النتائج إلى عدم قدرة محرك البحث (أين) على استرجاع المعلومات المناسبة، حيث لم تصل درجة البحث بأي من الأساليب الثلاثة إلى الرقم ٣، وهي الدرجة التي تشير إلى مستوى ارتباط المعلومات جزئياً باستفسار البحث.

ويؤكد تدني مستوى دقة الاسترجاع ما نتج عن تحليل جميع النتائج المسترجعة من استفسارات البحث كافة والبالغ عددها ١٨٠ استفساراً بمعدل

٦٠ استفساراً لكل مجموعة. حيث ظهر أن ٦٠ نتيجة فقط حصلت على الرمز (أ) وهو الذي يشير إلى تحقق الاستفادة الفعلية من المواقع فيما يرتبط باستفسارات البحث. وأن ما يزيد عن نصف النتائج وعددها ٩٢ نتيجة ليس لها اي ارتباط بالاستفسارات المطلوبة (انظر الشكل رقم ١).

الشكل رقم (١)

علاقة النتائج المسترجعة بالاستفسارات



محرك البحث (عربي):

يتضح من الجدول رقم (٢) أن النتائج المسترجعة من محرك البحث (عربي) باستخدام اللغة الطبيعية والربط البولياني منخفضة، كما هو الحال في محرك (أين). في حين حقق البحث عن طريق مطابقة العبارة درجة مرضية قدرها ٣,٧١ درجات، مما يعني أن استراتيجية البحث بهذه الطريقة أكثر كفاءة من الأساليب الأخرى، ويتحقق فيها الارتباط الجزئي باستفسارات البحث.

الجدول رقم (٢)

درجة التقويم لمحرك البحث عربي حسب نوع استراتيجية البحث

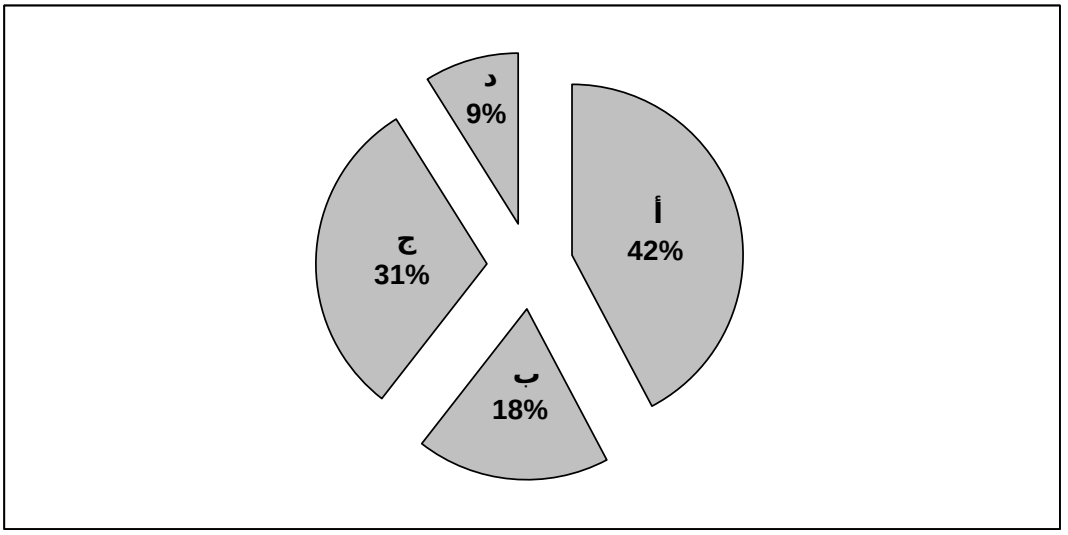
نوع الاستفسار	الدرجة
مطابقة العبارة	٣,٧١
اللغة الطبيعية	٢,٨٥
الربط البولياني	٢,٣٣

وبتحليل جميع نتائج البحث باستخدام الاستراتيجيات الثلاثة يتبين أن ٧٦ نتيجة بما نسبته ٤٢٪ حصلت على الرمز (أ) مما يدل على أنها ترتبط فعلياً باستفسارات البحث،

وأن ٣٣ نتيجة بما نسبته ٩٪ يتحقق فيها بعض الارتباط باستفسارات البحث. وعلى الرغم من أن هذه النتيجة أعلى من محرك البحث (أين) إلا أنها لا تزال منخفضة، وذلك لأن النتائج المتبقية ونسبتها ٤٩٪ لا ترتبط باستفسارات البحث؛ لأنها إما روابط لا تعمل أو مواقع تحتوي على معلومات لاعلاقة لها بموضوع البحث (انظر الشكل رقم ٢).

الشكل رقم (٢)

علاقة النتائج المسترجعة بالاستفسارات



محرك البحث ألتافيستا AltaVista :

حقق محرك البحث ألتافيستا AltaVista نتائج بحث مرضية بخلاف محركي البحث السابقين. حيث ارتفعت درجة كل الاستراتيجيات البحثية إلى مستوى عالٍ. فمن خلال قراءة الجدول رقم (٣) يلحظ أن استراتيجية البحث بمطابقة العبارة حقق ٤,٣٨ درجات، وهي درجة قريبة من المقياس الأعلى الذي يحدد درجة ارتباط النتائج باستفسارات البحث. يليه البحث باللغة الطبيعية وبالربط البولياني بدرجات متساوية حيث حقق كل منهما ٣,٧١ درجة. وتعني أن النتائج المسترجعة لها ارتباط جزئي باستفسار البحث، ويمكن أن تحقق بعض الفائدة للباحث. وتشير هذه النتائج مجتمعة إلى قدرة محرك البحث ألتافيستا AltaVista على استرجاع المواقع باستخدام الاستراتيجيات المختلفة، وعلى تفوقه على محركي البحث (أين) و(عربي) فيما يتعلق بمستوى دقة النتائج.

الجدول رقم (٣)

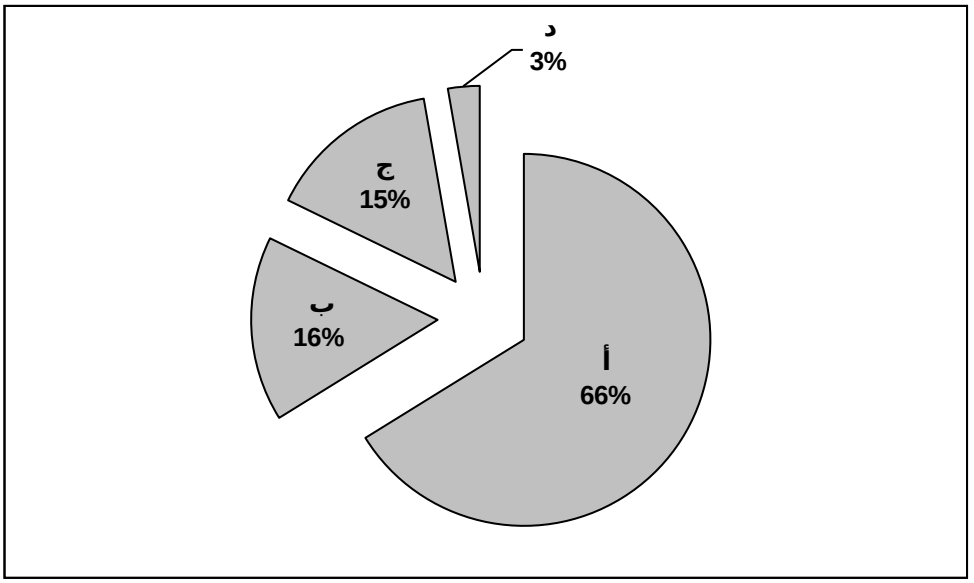
درجة التقويم لمحرك البحث التافيستا AltaVista حسب نوع استراتيجية البحث

نوع الاستفسار	الدرجة
مطابقة العبارة	٤,٣٨
اللغة الطبيعية	٣,٧١
الربط البولياني	٣,٧١

ويدعم هذه النتيجة أن ١١٩ موقعاً مسترجعاً بما نسبته ٦٦٪ من إجمالي الاستفسارات الموجهة لمحرك البحث التفيستاAltaVista حصلت على الرمز (أ) وهو الذي يشير كما وضح سابقاً إلى الارتباط الفعلي لمحتويات المواقع باستفسارات المطروحة (انظر الشكل رقم ٣).

الشكل رقم (٣)

علاقة النتائج المسترجعة بالاستفسارات



مقارنة محركات البحث الثلاثة:

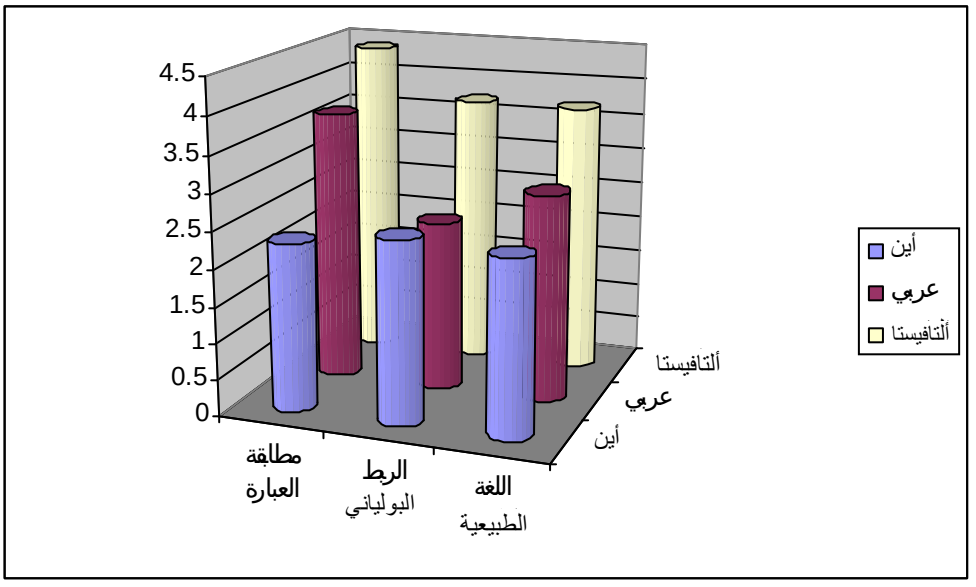
بعد التقويم السابق لكل محرك بحث على حدة، نعرض في هذا القسم نتائج التحليل المقارن بين المحركات الثلاثة للخروج بأحكام توضح قوة ودقة أداء محركات البحث العربية مقارنة بمحركات البحث الأجنبية ممثلة في ألتافيستا،

للتعرف إلى أفضلها في التعامل مع الاستراتيجيات المختلفة للبحث باللغة العربية.

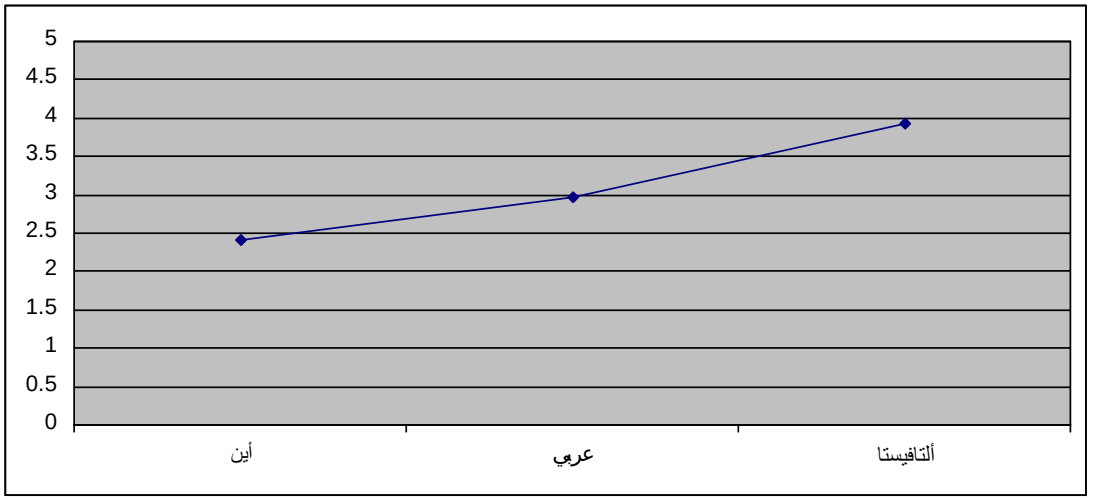
فمن خلال الشكل رقم (٤) يتبين أن محرك البحث ألتافيستا AltaVista تفوق على المحركين الآخرين في دقة استرجاع المعلومات وتوافقها مع الاستفسارات المطروحة باستخدام الأساليب المختلفة للبحث وبفارق مرتفع. كما أن مقارنة النتائج المسترجعة من استراتيجيات البحث مجتمعة وبشكل تراكمي تؤكد أيضًا حصول محرك ألتافيستا AltaVista على أعلى نسبة في دقة الاسترجاع وبفارق كبير عن نظيريه (انظر الشكل رقم ٥). حيث حصل المحرك ألتافيستا AltaVista على ٣,٩٤ درجة، وهو مستوى قريب من تحقيق الدرجة الأعلى في تطابق النتائج مع استفسارات البحث. يلي ذلك وبدرجات متقاربة المحرك (عربي)، وحصل على ٢,٩٦ درجة، والمحرك (أين) وحصل على ٢,٤١ درجة.

الشكل رقم (٤)

دقة استرجاع المعلومات وتوافقها مع الاستفسارات



الشكل رقم (٥)
النتيجة التراكمية لكل محرك بحث



مناقشة النتائج:

يبين التحليل السابق لنتائج الاستفسارات المعطاة لمحركات البحث المدروسة أن محركات البحث العربية لا تتوفر لديها القدرة الكافية في استرجاع المعلومات ومواقع الإنترنت المناسبة للباحث العربي. وتظهر معها الحاجة إلى بناء محركات بحث دقيقة تستطيع الربط بين طبيعة الاستفسار وبين النتائج المسترجعة لمد الباحث بالمعلومات التي يحتاجها وترتبط بشكل مباشر بسؤال بحثه. كما أن كثرة استرجاع الروابط المعطلة والمواقع المغلقة ضمن النتائج يعطي إشارة واضحة إلى أن محركات البحث العربية لا تقوم بتحديث كشافاتها التي يتم بناؤها عن طريق البواحث الإنترنتية. ويؤكد ذلك أن (٥٨) رابطاً وتشكل (١٦٪) من إجمالي النتائج المسترجعة من محركي (أين) و(عربي) والبالغ عددها (٣٦٠) رابطاً كانت معطلة. وهذه النسبة تعد عالية جداً إذا أخذنا في الحسبان أن هذه الروابط المعطلة هي من ضمن العشرة النتائج الأولى لكل استفسار موجه لمحركات البحث. وكما هو متعارف عليه فإن منهج استرجاع المواقع وترتيب عرضها للمستفسر يعتمد على الارتباط الموضوعي بسؤال البحث Relevance . ولو قارنا تلك النسبة مع الروابط المعطلة في

النتائج المسترجعة من محرك البحث ألتافيستا AltaVista لوجدنا فجوة كبيرة بين النسبتين لصالح ألتافيستا AltaVista ، حيث بلغت نسبة الروابط المعطلة ٠,٤٪ فقط.

وتلك النتائج تشير أيضاً إلى أن محركي البحث (أين) و (عربي) حققا نتائج متدنية جداً في الارتباط الموضوعي مع الاستفسارات المطروحة لهما لم تصل درجتها إلى المستوى الذي يشير حتى إلى الارتباط الجزئي بين استفسار البحث وموضوعات المواقع المسترجعة. حيث حقق المحرك (عربي) ٢,٩٦ درجة والمحرك (أين) ٢,٤١ درجة. كما أن هذه النتيجة تدعو إلى دراسة ميكانيكية البحث المتبعة في محركات البحث الأجنبية ومحاولة تعميمها على محركات البحث العربية، لتكون منطلقاً لبناء محركات عربية متكاملة قادرة على التحليل الموضوعي للمواقع العربية على الإنترنت، واسترجاع نتائج مرضية للباحثين في المواقع العربية.

وبالإضافة إلى الحاجة إلى تطوير منهج التحليل الموضوعي لمواقع الإنترنت العربية، فإن الحاجة أيضًا تدعو إلى أن تنظر محركات البحث إلى نوعية هذه المواقع والمسئولية التبعية للجهات المشرفة عليها. حيث تبين من تحليل النتائج المسترجعة والتي ترتبط موضوعياً باستفسارات البحث أن نسبة كبيرة منها منتديات حوارية، وليست مواقع لمؤسسات علمية أو جهات رسمية يمكن الثقة فيما تضعه من معلومات على موقعها. كما أن بعض المعلومات المدرجة في المنتديات هي معلومات منقولة من مواقع رسمية أو مجلات علمية ومع ذلك لا يتم إدراج المواقع الأصلية التي نقلت منها هذه المعلومات ضمن النتائج المسترجعة.

الخاتمة:

هدفت هذه الدراسة إلى قياس مدى كفاءة أداء محركات البحث العربية في استرجاع المعلومات والمواقع العربية على شبكة الإنترنت عن طريق تقييم نتائج البحث لمجموعة من الاستفسارات لمحركي بحث عربيين هما (أين) و (عربي)، ثم مقارنة تلك النتائج مع محرك البحث الأجنبي ألتافيستا AltaVista. وقد توصلت الدراسة إلى أن محركات البحث العربية لا تزال ضعيفة ودون المستوى المطلوب في استرجاع المعلومات العربية، ولم تحقق مستوى ارتباط عاليًا بين الاستفسارات المطروحة وبين النتائج المسترجعة. وبناء عليه فقد خرجت الدراسة بمجموعة من التوصيات أبرزها:

١- ضرورة تطوير آليات البحث المتبعة في محركات البحث العربية؛ لتتوافق وطبيعة اللغة

العربية، ولتكون لديها القدرة في تحليل المحتوى الموضوعي وتكشيفه للمواقع العربية.

٢- الحاجة إلى تطوير نظام استرجاع عربي متقدم وقادر على فهم النصوص العربية وبنائها اللغوي.

٣- ضرورة تحديث كشافات محركات البحث العربية عن طريق مسح المواقع العربية على الإنترنت بشكل مستمر وعلى فترات قصيرة، لتجنب ظهور المواقع المغلقة والروابط المعطلة ضمن نتائج البحث.

٤- أظهرت الدراسة ضعف محركات البحث العربية، وبالتالي توصي الدراسة الباحثين العرب بعدم الاعتماد عليها في الوقت الحاضر لاسترجاع المعلومات من الإنترنت، والاستعانة بدلاً عنها بمحركات البحث الأجنبية الشهيرة التي تدعم اللغة العربية مثل جوجل Google وألتافيستا AltaVista.

٥- من الأفضل أن يحدث تغيير في خوارزميات الفرز والترتيب لنتائج البحث، بحيث تستطيع المحركات العربية التفريق بين المصادر الأساسية للمعلومات وبين المصادر التي تقوم بنقل تلك المعلومات إليها، مثل المنتديات الحوارية. ويتم بناء على ذلك إدراج المصادر الأصلية للمعلومات في مستويات متقدمة ضمن نتائج البحث.

٦- أن تتولى منظمات إقليمية ومؤسسات مهنية متخصصة تصميم محركات البحث العربية، بدلاً من الاعتماد على الجهود الفردية. وذلك لأن الجهد الفردي غالباً ما ينقصه الدعم الكافي ويكون مصيره الفشل في نهاية المطاف.

المراجع

- خالد عبدالفتاح محمد. مصداقية محركات البحث في تكشيف وترتيب الصفحات المكررة المتاحة باللغة العربية. مجلة المكتبات والمعلومات العربية. س٢٧. ع٢. أبريل ٢٠٠٧م.
- عامرة حقي الفرغولي. تطبيق الطرق الإحصائية في تحليل محتوى النصوص بلغتها العربية آلياً. العربية ٣٠٠٠. العدد ١. ٢٠٠٢م.
- نادية مصطفى العيدروس أحمد. دراسة مقارنة لتقييم أثر محركات البحث العربية والعالمية الداعمة للغة العربية في استرجاع المعلومات. الخرطوم: المركز القومي للبحوث، ٢٠٠٦م.
- Bar-Ilan, J., Levene, M., & Mat-Hassan, M. (2006) Methods for evaluating dynamic changes in search engine rankings: a case study. *Journal of Documentation, 62*(6), 708-729.
- Feng, J., Bhargava, H, & Pennock, D. (2007). Implementing sponsored search in Web search sngines: computational evaluation of alternative mechanisms. *Informs Journal on Computing, 19*(1), 134-148.

search engines. NTUA, Retrieved from http://piper.ntua.gr/reports/searcheng/doc0000.htm

- Rainie, L., & Shermak, J. (2005, November 20). *Search engine use shoots up in the past year and edges towards email as the primary internet application*. Washington, D.C.: Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/pdfs/PIP_SearchData_1105.pdf

- Shultz, M. (2007). Comparing test searches in PubMed and Google Scholar. *Journal of the Medical Library Association, 95*(4), 442-445.

- Vaughan, L. (2004). New measurements for search engine evaluation proposed and tested. *Information Processing & Management,* 40(4), 677-691.

- Griesbaum, J. (2004). Evaluation of three German search engines: AltaVista.de, Google.de and Lycos.de. *Information Research, 9*(4). Retrieved from http://eprints.rclis.org/archive/00002772/01/paper189.html

- Isfandyari Moghaddam, A. (2007). Web metasearch engines: a comparative study on search capabilities using an evaluation checklist. *Online Information Review, 31*(3), 300-309.

- Jacobs. J, Rubens. M. (2003). An online relevancy tuning algorithm for search engines. Retrieved from http://www.jaredjacobs.com/stanford/cs276/writeup.pdf.

- Lewandowski, D. (2005). Web searching, search engines and information retrieval. *Information Services & Use, 25*(3), 137–147.

- Nasios, Y., Korinthios, G., Despotopoulos, Y. (1998). Evaluation of

* * * *