

## تغطية صفحات المعلومات العربية فى أدوات البحث على الشبكة العنكبوتية العالمية

د. أمجد عبد الهادى الجوهري

مدرس علم المعلومات كلية الآداب

جامعة المنيا

aaagohary@yahoo.com

د. خالد عبد الفتاح محمد

مدرس علم المعلومات كلية الآداب

جامعة المنيا

Khaledma@hotmail.com

### ١- تمهيد:

على الرغم من أن هذه المشاركة ما زالت محدودة وعلى الرغم من التساؤلات العديدة المطروحة حول كفاءة وجودة تلك المصادر وإمكانية الاعتماد عليها فى الدراسات والبحوث، إلا أن تلك الشبكة أصبحت منفذ لا يمكن الاستغناء عنه فى عمليات التواصل والنشر والبحث.

ومن الأمور الجديرة بالاهتمام والفحص الزيادة المطردة (انظر ملحق ١) فى عدد مستخدمي شبكة الإنترنت، حيث بلغ عدد المستخدمين فى عام ١٩٩٨ حوالى ١٥١ مليون مستخدم. وقد وصل هذا العدد مع بداية عام ٢٠٠٤ إلى ٤٩٥ مليون مستخدم. وقد قدرت الإحصائيات أيضاً أن يصل عدد المستخدمين إلى ٩٦٤, ٢٨٩, ٧٠١ مستخدم خلال عام ٢٠٠٥. وقد بلغت الزيادة المطردة فى عدد مستخدمي الشبكة خلال الخمس سنوات (٢٠٠٥-٢٠٠٠) ما يعادل ١, ١٦٧, ١ Internet %

تعتبر شبكة الإنترنت بصفة عامة والشبكة العنكبوتية العالمية (World Wide Web (WWW بصفة خاصة من أهم الملامح المميزة لمجتمع المعلومات المعاصر. وذلك لما تتمتع به من ضخامة وتنوع فى مصادر المعلومات المتاحة من خلالها وما تقدمه من خدمات لجمهور المستخدمين فى العالم بصفة عامة وفى مجتمعنا العربى على وجه الخصوص. وقد أتاحت تلك الشبكة أفقاً جديدة للباحث والقارئ العربى فى الانفتاح والتواصل مع المجتمعات الأخرى بحيث أصبح من الممكن الحصول على كثير من مصادر المعلومات طريق خدمات البحث والاسترجاع المتاحة على الشبكة العنكبوتية. كما أصبح من الممكن لأى باحث أن ينشر بحثه أو دراسته على شبكة الإنترنت مما يساهم فى المشاركة الفعالة فى الإنتاج الفكرى العالمى،

الصفحات المتاحة والثابتة قد تزايد من ٣٠٠ مليون صفحة عام ١٩٩٨ إلى ٢٠ بليون صفحة عام ٢٠٠٢، بينما يقدر عدد الصفحات المتاحة حتى ديسمبر ٢٠٠٤ على الويب ٢٠٠ بليون صفحة (Gromov, Gregory, 2004). أما عن حجم المشاركة العربية في بناء صفحات الويب فقد بلغت ١٨ ألف صفحة عام ١٩٩٨، وتزايدت بسرعة كبيرة خلال بداية الألفية الجديدة حيث قدرت مجموعة مدار للبحث أنها من المحتمل أن تبلغ أكثر من ٨ مليون صفحة في بداية عام ٢٠٠٥ (Madar Research Group, 2004). ونظراً لضخامة وتنوع مصادر المعلومات المتاحة على الشبكة العنكبوتية أصبح من الصعب الوصول إلى تلك المصادر دون وجود أدوات تيسر عمليات البحث والاسترجاع، وتنقسم تلك الأدوات إلى نوعين أساسيين هما:

#### \* أدلة البحث Directories :

وهي عبارة عن فهراس مصنفة تحت رؤوس موضوعات عريضة وكل رأس موضوع يتشعب إلى مجموعة من الرؤوس الفرعية ولعل أبرز مثال لها يتمثل في دليل البحث Yahoo الذي تحول في الآونة الأخيرة إلى محرك بحث حيث قام بشراء أداة البحث من شركة Alltheweb ويتطور بسرعة كبيرة لكي يمثل بوابة رئيسية من بوابات الويب<sup>(١)</sup>، (Yahoo, 2005). كما توجد العديد من أدلة البحث العربية التي تكشف صفحات

Usage Statistics: Usage and Poupulation Statistics, 2005, Sullivan, 2005b كما تشير الإحصائيات أيضاً إلى التزايد المطرد في عدد مستخدمي الشبكة في المنطقة العربية حيث بلغ عام ١٩٩٨ سبعمائة ألف مستخدم، تزايد إلى ١,٥ مليون مستخدم عام ١٩٩٩، ارتفع هذا العدد إلى ١٥ مليون مستخدم في نهاية عام ٢٠٠٢، وبلغ حوالي ٢٥ مليون مستخدم عام ٢٠٠٥ منهم ٦,٥ مليون مستخدم من جمهورية مصر العربية وحدها (Madar Research Group, 2004). وتسير الزيادة المطردة في عدد المستخدمين جنباً إلى جنب مع الزيادة المطردة في عدد الصفحات المتاحة على الشبكة العنكبوتية العالمية، وعلى الرغم من هذه الزيادة إلا أنه من الصعب الحصول على إحصائية دقيقة عن عدد الصفحات المتاحة على الشبكة لسببين أساسيين هما (Notess, 1997):

- أن معظم التقديرات عادة ما تعتمد على قياس عدد الصفحات الثابتة Static Pages وتتجاهل الصفحات الديناميكية والتفاعلية Dynamic & Interactive Pages وهي تمثل نسبة غير قليلة ومن الصعب الوصول إليها.
  - التزايد الكبير في عدد الصفحات كل ثانية في جميع أنحاء العالم مما يجعل من الصعب وبل ومن المستحيل متابعتها بصورة دقيقة.
- ومع ذلك فإن التقديرات تشير إلى أن عدد

(١) يشير مصطلح البوابات إلى مواقع ويب تقدم مجموعة متنوعة من خدمات المعلومات للمستخدمين منها بغرض جذب أكبر عدد ممكن من المستخدمين وتنقسم إلى بوابات عامة يطلق عليها البوابات الأفقية وأخرى متخصصة يطلق عليها البوابات الرأسية.

المعلومات العربية ومن أمثلتها: فارس نت<sup>(٢)</sup>، ردادى<sup>(٣)</sup>، بوابة العرب<sup>(٤)</sup>.

تعتمد أدلة البحث على أساليب التشفير اليدوى التقليدية، حيث يقوم مجموعة من المتخصصين بتصفح الويب بشكل دورى لتجميع قوائم بصفحات المعلومات التى تظهر على الشبكة ثم يتم بث هذه القوائم إلى مجموعة من المكشفين المتخصصين. يقوم المكشوفون بتصفح الصفحات الجديدة وتحديد رؤوس الموضوعات العريضة والفرعية التى تدرج تحتها تلك الصفحات وإضافة عنوان الصفحة إلى الدليل وربطه بموقع الصفحة على الشبكة العنكبوتية.

#### \* محركات البحث Search Engines:

هى أدوات تعتمد فى الأساس على التجميع والتشفير الآلى لصفحات المعلومات المتوافرة على الشبكة العنكبوتية العالمية. ومن أبرز محركات البحث العالمية المتاحة فى الوقت الحالى .GOOGEL, AltaVista, LYCOS, Hotpot. أما على المستوى العربى، فعلى الرغم من تنوع أدلة البحث إلا أن الفحص الدقيق لها أثبت أنه لا يوجد سوى محرك بحث عربى واحد هو أداة البحث أين (محمد فتحى عبد الهادى، ٢٠٠١؛ داليا نصار، ٢٠٠٤). فمعظم أدوات البحث العربية لا تعدو أن تكون مجرد أدلة بحث تمتلك إمكانيات البحث فى الأدلة Directory Search أو أن البحث فيها مدعم بإمكانيات البحث فى أحد المحركات العالمية

Search Powered By وتشير إحصائيات محرك البحث أين إلى أنه يغطى ٢١٣٠٧٠٨ صفحة معلومات وهو بالطبع عدد محدود جداً قياساً على ما تغطيه محركات البحث العالمية (أين، ٢٠٠٥).

وتتفاوت محركات البحث من حيث تغطيتها لمصادر المعلومات المتاحة على الشبكة العنكبوتية. ففى دراسة لحجم التغطية لصفحات المعلومات فى محركات البحث العالمية أوضح لورانس وجيل (Lawrence and Giles, 2002) أن هناك ٨٠٠ مليون صفحة ويب متاحة للجمهور العام Publicly ويمكن تكشيفها من خلال أى محرك بحث، وأن ٤٢٪ من هذه الصفحات مكشوفة من خلال ٦ محركات بحث عالمية، وأن أى من هذه المحركات لا يغطى منفرداً أكثر من ١٠٪ من صفحات الويب القابلة للتكشيف Indexable Web.

وتتميز محركات البحث بقدرتها على التجميع الآلى لصفحات الويب من خلال الاعتماد على زواحف أو العناكب Crawlers or Spiders تتولى مسح الشبكة العنكبوتية للتعرف على الصفحات الجديدة وتجميعها. كما تتميز أيضاً باستخدام تقنيات التشفير الآلى للنصوص والروابط الفائقة. وقد أدى ذلك إلى أن حجم التغطية فى محركات البحث يفوق بدرجة كبيرة معدلات التغطية فى أدلة البحث التى تعتمد بدرجة كبيرة على التجميع والتشفير اليدوى. وأثبت التحليل

www.fares.net

www.radai.com

www.arabsgate.com

(٢)

(٣)

(٤)

الدراسة إلى ثلاثة فئات أساسية هي: الإنترنت واسترجاع المعلومات، الدراسات الكلاسيكية<sup>(٦)</sup> في التغطية، دراسات التغطية في محركات وأدلة البحث. وقد قام الباحثان باختيار مجموعة من الدراسات تعكس اتجاهات الإنتاج الفكرى فى كل فئة من هذه الفئات. وفيما يلي عرض لمجموعة الدراسات السابقة المرتبطة بكل فئة:

### ★ الإنترنت واسترجاع المعلومات:

أدى تطور تقنيات الشبكة العنكبوتية منذ بداية التسعينيات إلى زيادة اهتمام الباحثين فى مجال المعلومات بدراسة موضوعات متفرقة مرتبطة بتلك التقنيات، وما لها من آثار على تيسير سبل الإفادة من مصادر المعلومات العامة والمتخصصة المتاحة على تلك الشبكة. ووفقاً لدراسة لاين (Line, 1998) أصبحت الإنترنت تمثل أبرز المجالات البحثية فى تخصص علم المعلومات فى الوقت الراهن. وتعد أدوات الوصول إلى المعلومات على الشبكة العنكبوتية ومن أبرزها أدلة ومحركات وما وراء البحث من أكثر الموضوعات التى تحظى بعناية الباحثين فى مجال علم المعلومات. وقد تنوعت دراسات أدوات الوصول إلى المعلومات على الشبكة العنكبوتية ما بين وصفية ومقارنة وتجريبية (Chowdhury, 1999). وعلى الرغم من هذا

الديق لعدد من أدلة البحث العالمية أنها لا تغطى أكثر من 7.5 من الصفحات القابلة للتكشيف (Lawrence and Giles, 1999). ذلك على الرغم من أن عمليات التكشيف فى أدلة البحث عادة ما تكون أكثر دقة نظراً لأنها تعتمد على الاختيار الدقيق للصفحات التى يتم تغطيتها فى الأدلة إلى جانب تكشيف الصفحات تحت رؤوس موضوعات يتم اختيارها بعناية ودقة كبيرة.

من ثم فدراسة تغطية محركات وأدلة البحث لصفحات الويب ضرورة ملحة، حيث أن هذه النوعية من الدراسات تجيب على سؤال هام هو: ما هى أكثر أدوات البحث شمولاً فى تكشيف واسترجاع صفحات الويب<sup>(٥)</sup>؟ وتركز هذه الدراسة بصفة خاصة على الأدوات العالمية دون الأدوات العربية نظراً لأن أدوات البحث العربية ما زالت محدودة وانتقائية ومتحيزة فى تغطيتها إلى حد كبير. وهو ما اتضح من خلال تجربة رائدة قام بها الباحثان قبل الشروع فى هذه الدراسة (انظر ص ١٥ - ١٦)، لذلك تحول تركيز الباحثان إلى الأدوات العالمية وتم الاعتماد على الأدوات العربية فى تجميع عينة الدراسة.

### ٣ - الدراسات السابقة:

تم تقسيم الإنتاج الفكرى المرتبط بموضوع

(٥) لم يتطرق الباحثان إلى ما وراء المحركات Meta Search Engines وبوابات الويب Web Portals حيث أنها أدوات تساعد على عمليات البحث والاسترجاع ولكنها لا تقوم بتجميع وتكشيف صفحات الويب ولا تقوم ببناء قواعد بيانات خاصة بها بينما تعتمد على قواعد البيانات التى تقوم ببناءها محركات وأدلة البحث.

(٦) مجموعة الدراسات التى ظهرت مع بدايات تقييم نظم استرجاع المعلومات التى وضعت معايير ومقاييس تقييم الأداء فى تلك النظم.

الاهتمام إلا أنه ما زال هناك نقص شديد فى الدراسات العربية فى مجال علم المعلومات المتعلقة بأدوات الاسترجاع فى بيئة الإنترنت والشبكة العنكبوتية بشكل عام. وبمراجعة الإنتاج الفكرى العربى لم يجد الباحثان أى دراسة تتناول من قريب أو بعيد موضوع تغطية مصادر المعلومات العربية فى محركات وأدلة البحث العالمية أو العربية على الرغم من وجود عدد من الدراسات التى تناولت موضوع التغطية سواء فى قواعد البيانات الببليوجرافية أو مؤسسات المعلومات. وتتنحصر معظم الكتابات العربية فى هذا الموضوع فى إطار الكتابات التعريفية فى نطاق الدوريات وأوراق المؤتمرات. ومن أمثلة تلك الدراسات التعريفية دراسة (خالد رياض، ١٩٩٩) الذى قدم فيها لأدوات البحث على الإنترنت. بعض الدراسات اهتمت بالجانب الوصفى لبعض فئات المواقع، ومن أبرز تلك الدراسات دراسة لمواقع الثقافة العربية المتاحة عبر الإنترنت سواء تلك المتوافرة من خلال خوادم داخل الوطن العربى أو خارجه (سعد الهجرسى، ٢٠٠٢). وقد أدى تطور القدرات الاسترجاعية لمحركات البحث وزيادة اهتمام المستفيدين باستخدامها إلى ظهور بعض الدراسات العربية التجريبية فى السنوات الأخيرة ومن أمثلتها دراسة زين عبد الهادى (٢٠٠٢) والتى ركز فيها على تقييم خصائص عدد من محركات البحث بشكل تجريبى لقياس إمكانيات البحث فى تلك المحركات وزمن الاستجابة بالإضافة إلى مدى التداخل بين النتائج المسترجعة. وقد خلصت الدراسة إلى مجموعة من العوامل المؤثرة فى نتائج البحث ومن أهمها: مدى تنقيح محركات البحث لكشافاتها ومدى تغطية تلك الكشافات للمواقع.

وفى نفس الإطار تناولت دراسة الماجستير وصف وتقييم محركات البحث العربية على الإنترنت ومدى قدرة تلك المحركات على دعم البحث باللغة العربية من خلال استخدام المنهج التجريبى (داليا نصار، ٢٠٠٤). وقد أرجعت الباحثة بداية تلك المحركات إلى عام ١٩٩٥ حيث ظهور دليل البحث أين. كما أوضحت أن من بين الأدوات العربية العديدة التى تستخدم فى البحث والاسترجاع من الشبكة العنكبوتية، أن محرك البحث أين هو المحرك الوحيد الذى تنطبق عليه الخصائص العامة التى تميز المحركات عن أدلة البحث. وأن كل هذه الأدوات لا تعدو عن كونها أدلة موضوعية للشبكة العنكبوتية. وقد عرض محمد فتحى عبد الهادى (٢٠٠١) فى دراسة وصفية تعريفية لخدمات المكتبات والمعلومات المتاحة من خلال شبكة الإنترنت. قدم فيها تعريفاً إجرائياً لمحركات البحث كما أشار إلى مجموعة من المحركات العربية التى كانت ما زالت فى وقت إعداد الدراسة تجارب محدودة إلى حد كبير. وعرض ضياء الدين (٢٠٠٤) بالمقارنة بين واجهات التعامل فى مجموعة من محركات البحث العالمية بالإضافة إلى المقارنة بين كفاءة تلك الواجهات. كما قدم خالد عبد الفتاح (٢٠٠٥) لموضوع ما وراء المحركات Meta Search Engines حيث عرض لأسلوب بناء وطريقة عمل تلك الأدوات التى تستخدم فى بحث أكثر من محرك فى نفس الوقت. كما قارن أيضاً فى دراسته للدكتوراه بين مجموعة من أساليب دمج وفرز البيانات فى ما وراء المحركات لتحديد أفضل أساليب الدمج والفرز التى تصلح للعمل فى بيئة الويب والتى تتميز عن غيرها

من بيئات استرجاع المعلومات أنها بيئة تعمل على الهواء<sup>(٧)</sup> (Mohamed, 2005) On the Fly .

### ★ الدراسات الكلاسيكية في التغطية:

يعتبر موضوع التغطية من الموضوعات التي حظيت بعناية واهتمام خاص من جانب المتخصصين في مجال المكتبات وعلم المعلومات، وبصفة خاصة المهتمين بأدوات ونظم استرجاع المعلومات. يرجع سبب الاهتمام بهذا الموضوع إلى أنه يجيب عن تساؤل هام وضروري لمستخدمي المكتبات وأدوات البحث والاسترجاع وهو: ما هي الأداة الأكثر شمولاً أى التي تغطي أكبر قدر ممكن من احتياجات المستفيدين. ويشير كلاً من لانكستر وورنر (Lancaster, Warner, 1993, p 116) إلى أن تقييم التغطية في قواعد ومراسد البيانات البليوجرافية يشبه إلى حد كبير تقييم مدى شمول واكتمال مجموعات المكتبات. ويقصد بالتغطية في دراسات استرجاع المعلومات «مدى شمول المجموعات التي يكشفها نظام استرجاع المعلومات من الناحية الموضوعية والشكلية والجغرافية والزمنية». وتجدر الإشارة هنا إلى أن الشمول مسألة نسبية تتحدد وفقاً لسياسة القائمين على نظام استرجاع المعلومات.

وقد ارتبطت التغطية الموضوعية في معظم الدراسات الكلاسيكية لاسترجاع المعلومات

بموضوعات التداخل والتكرار والتفرد في التغطية حتى أصبحت من المقاييس الأساسية التي يعتمد عليها في تقييم نظم استرجاع المعلومات. ويشير لانكستر إلى ثلاث طرق أساسية يمكن الاعتماد عليها في قياس التغطية الموضوعية والتداخل والتكرار لنظم استرجاع المعلومات. ويمكن تلخيص هذه الطرق فيما يلي: (Lancaster, 1998, p. 120 - 124)

- الاعتماد على القوائم البليوجرافية الشاملة (وهي الطريقة المستخدمة في هذه الدراسة).
- الاعتماد على الاستشهادات المرجعية لمجموعة من الدراسات الحديثة.
- الاعتماد على عينة ممثلة من الدوريات البورية والثانوية في التخصص الموضوعي.

الإنتاج الفكري حول موضوع التغطية ملئ بالنماذج التي تقارن بين مجموعة من قواعد البيانات في مجال موضوعي واحد أو مجموعة من المجالات المرتبطة. ونكتفى هنا بذكر مجموعة من الأمثلة التي تناولت هذا الموضوع. فقد قام ريد وزملاءه (Read, E., E. J., et. Al., 2000) بالمقارنة بين ثلاث قواعد بيانات متخصصة في مجال المكتبات والمعلومات هي LISA, ISA, Library Literature<sup>(٨)</sup> لتحديد أكثرهما شمولاً للإنتاج الفكري المتخصص في هذه المجالات. كما

(٧) البيانات التي تعمل على الهواء تتميز بأن عمليات الكشف والاسترجاع مستمرة في قاعدة البيانات حيث يتم إضافة صفحات جديدة إلى قاعدة البيانات كل ثانية دون الحاجة إلى فصل قاعدة البيانات وتعديل محتوياتها بالتالي يمكن أن تختلف نتائج البحث في محركات البحث كل ثانية.

(٨) LISA (Library and Information Science Abstract) ISA (Information Science Abstract).

قام بيرير بتقييم شمول التغطية للدوريات الإلكترونية في قواعد البيانات المقتناة في مكتبة جامعة هاواي لتحديد مدى تغطية كل عناوين الدوريات الإلكترونية في تلك القواعد ومدى الشمول في تكشيف محتوياتها (Brier, D. J., et. al., 1999).

أما على المستوى العربي فالدراسات في هذا المجال محدودة نظراً لقصور نظام الضبط البليوجرافي العربي مما نتج عنه ندرة في نظم استرجاع المعلومات العربية. ومن الدراسات التي تناولت موضوع التغطية على المستوى العربي تقييم مدى شمول التغطية في قواعد البيانات القطاعية للإنتاج الفكرى المصرى فى العلوم والتكنولوجيا. وقد اعتمدت هذه الدراسة على تجميع قائمة شاملة بالدوريات المصرية المتخصصة فى كل قطاعات العلوم والتكنولوجيا التى تدعى مراكز المعلومات القطاعية فى مصر تغطيتها وتكشيفها. وقد قام الباحث فى هذه الدراسة بمقارنة عينة من مقالات الدوريات المنشورة خلال فترة ثلاث سنوات وفقاً للقطاعات الموضوعية التى تغطيها كل قاعدة لتحديد مدى شمول التغطية بالإضافة إلى التداخل والتكرار بين قواعد البيانات (خالد عبد الفتاح محمد، ١٩٩٩).

#### ★ التغطية فى محركات وادلة البحث:

منذ ظهور أول محرك بحث على الويب عام

١٩٩٤ ودراسات التغطية فى محركات البحث فى زيادة مستمرة، وذلك بغرض الإجابة على سؤال هام وملح حول أكثر المحركات تغطية لصفحات الويب من أجل اختيار المحرك المناسب لبحث الشبكة العنكبوتية (Lenssen, 2004; Unerwood, 2005). وقد قام كلا من بهارات وبرودر (Bharat, K., & Broder, A. 1998) بتطوير طريقة لحساب التغطية فى محركات البحث بدلاً من الاعتماد على القيم المحدودة التى تنشرها محركات البحث حول عدد الصفحات التى تغطيها قواعد بياناتها. وقد وجد الباحثان أنه من بين أكثر أربعة محركات بحث، أن التغطية لمصادر المعلومات المتاحة على الشبكة العنكبوتية تتراوح ما بين ١٧ - ٤٧٪. كما أوضح لورانس وجيل (Lawrence, S. & Giles, C. L., 1999) أنه من بين أكبر ٦ محركات بحث لا يوجد أى من هذه المحركات يغطى أكثر من ثلث الصفحات المتاحة للتكشيف<sup>(٩)</sup> Indexable Web وأن أقل المحركات تغطية لا يغطى أكثر من ٣٪ من الصفحات المتاحة للتكشيف. وفى دراسة أخرى أعدها لورانس وجيل (Lawrence, S. & Giles, C. L. 2002) أوضحوا فيها أن التغطية قد انخفضت مع النمو المستمر فى عدد الصفحات وعدم قدرة محركات البحث على ملاحقة هذا النمو. وقد كشفت الدراسة أن أكبر محركات البحث من حيث التغطية

(٩) الصفحات المتاحة للتكشيف هى الصفحات الثابتة Static Pages وهى الصفحات التى لا تحتاج إلى أى شكل من أشكال تحديد الهوية لاستخدامها حيث تكون متاحة للجمهور العام ولها عناوين ثابتة على الويب. أما الصفحات الغير متاحة للتكشيف هى الصفحات الديناميكية والتفاعلية التى تحتاج إلى تحديد هوية أو يتم تخزينها فى قواعد بيانات خاصة بالتالى لا يمكن لزواحف محركات البحث اصطفاها وتكشيفها.

وقد قام نوتيس (Notess, 1997) بقياس حجم الصفحات والمواقع المتاحة على شبكة الإنترنت بالاعتماد على تقدير حجم الصفحات المكشوفة في ثماني محركات بحث عالمية. كما قام كلا من هينزينجر وزملاؤه (Henzinger et al. 1999) باختبار مدى تكشيف صفحة معينة في عدد من محركات البحث وذلك بالاعتماد على أسلوب الواقعة الحاسمة Critical Incident لتقييم شمول التغطية في محركات البحث، وذلك من خلال تتبع الروابط الفائقة للصفحة للتعرف على مدى تكشيف الصفحة الرئيسية والصفحات المرتبطة بها في كل المحركات محل الدراسة. كما قام كلا من فوغان وثيلوال (Vaughan & Thelwall, 2004) بقياس التحيز في تغطية محركات البحث العالمية Search Engines Coverage Bias. وذلك من خلال المقارنة بين مدى تغطية الصفحات التجارية والحكومية المتاحة على خوادم ٤٢ دولة. وأوضحت الدراسة وجود درجة كبيرة من الاختلاف في تغطية تلك المحركات على سبيل المثال وجد أن Alta Vista يغطي ٨٢٪ من المواقع الفرنسية بينما يغطي فقط ٣٦٪ من المواقع المصرية. وقد أثبتت الدراسة تحيز محركات البحث للصفحات المتاحة على خوادم في الولايات المتحدة. كما اكتشف كلا من موشيويتز وكاوجشي (Mowshowitz & Kawaguchi, 2002) طريقة جديدة لقياس التحيز في التغطية Coverage Biasness من خلال اختبار النتائج التي يسترجعها أحد محركات البحث ومقارنتها بالنتائج التي تسترجعها مجموعة من المحركات مجتمعه. كما قام مقدار بقياس مدى تعرف محركين بحث

لا يغطي أكثر من ١٦٪ من الصفحات القابلة للتكشيف. كما أوضحنا أن هذا التناقص المستمر في حجم التغطية يرجع إلى فاعلية وعائد التكلفة والقيود التكنولوجية التي تفرض على سعة عمليات التكشيف والاسترجاع والقيود التي تفرض على سعة الشبكة الداخلية لمحركات البحث Search Engine Internal Network Bandwidth. هذا وإن كنا نتفق مع كل هذه الأسباب التي طرحت، فإننا نضيف إلى ذلك أن تركيز محركات البحث على صفحات المعلومات التي تنتج وتتاح من خوادم داخل الولايات المتحدة وأوروبا - نظراً لسهولة التعرف عليها - مما يؤدي إلى تناقص التغطية مع الزيادة المطردة في عدد الصفحات التي تنتج وتتاح من دول أخرى على الشبكة العنكبوتية ودخول دول ومجالات موضوعية لم يكن لها أي تواجد في المراحل الأولى من نشأة الشبكة العنكبوتية، كل ذلك أدى إلى تناقص التغطية بدرجة كبيرة عندما يتم قياس التغطية في نطاقات أوسع Larger Scale وهو ما أوضحته مجموعة من الدراسات مثل (Notess, 2005; Vaughan & Thelwall, 2004).

ويجب أن نشير في هذا السياق إلى أنه لا يمكن الاعتماد على النتائج السابقة نظراً للطبيعة المتغيرة للويب، إلا أنه توجد العديد من المواقع التي توفر بيانات أكثر حداثة عن حجم التغطية في محركات البحث مثل:

<http://www.searchengineswatch.com>

<http://www.searchenginesshowdown.com.stats/size.shtml>

<http://bigsearchengineindex.com>



هاوكنج وزملاؤه (Hawking, et. al, 2001) لم يجدوا أى علاقة ارتباط Correlation بين التغطية ودقة الاسترجاع. إلا أنهم أشاروا إلى أن التغطية عامل فى غاية الأهمية لأنواع معينة من الاستفسارات. كما أوضح عرض الإنتاج الفكرى أن موضوع التغطية من الموضوعات التى حظيت بعناية واهتمام خاص فى مجال المكتبات والمعلومات، كما أنه لا توجد أية دراسة عربية تناولت موضوع تغطية صفحات المعلومات العربية فى أدوات البحث على الشبكة العنكبوتية مما يعكس ضرورة وأهمية هذه الدراسة.

### ٣ - موضوع الدراسة:

توجد العديد من الأدوات التى تسعى إلى تكشيف وإتاحة صفحات المعلومات العربية للبحث والاسترجاع منها محركات وأدلة البحث العالمية ومحركات وأدلة البحث والصفحات الصفراء العربية<sup>(١٠)</sup>. إلا أن أدوات البحث العربية ما زالت محدودة فى تغطيتها وقدرتها على التجميع والبحث (داليا نصار، ٢٠٠٤). من ثم تعد أدوات البحث العالمية من أدلة ومحركات بحث هى المصدر الأساسى لاسترجاع الصفحات العربية. لذلك تسعى هذه الدراسة إلى تحليل حجم تغطية صفحات المعلومات العربية فى محركات وأدلة البحث العالمية وذلك من خلال المقارنة بين حجم التغطية فى تلك الأدوات. هذا بالإضافة إلى دراسة حجم التغطية لثلاثة مجالات أساسية هى الصحة والحاسب الآلى والإسلاميات، إلى جانب قياس معدلات التداخل

مختلفين على حروف اللغة العربية وقدرتهما على تكشيف واسترجاع المواد العربية، وبالتالي تحقيق أعلى مقاييس التغطية للمواقع العربية (Moukdad, 2002). وقد أوضح موقع الويب <http://www.searchengienwatch.com> فى ديسمبر من عام ٢٠٠٤ أن GOOGLE يعد أكبر محركات البحث من حيث التغطية ويبلغ حجم قاعدة بياناته ٢ بليون صفحة. وقد بلغ حجم قواعد البيانات لعدد من محركات البحث الشهيرة مثل Altavista, Alltheweb and Yahoo ١٠٠ مليون إلى نصف بليون صفحة. ويرجع تفوق محرك البحث Google إلى أسلوب التكشيف الذى يستخدمه حيث يعتمد على تحليل روابط الويب Web Hyperlinks Analysis، مما جعله يدعى - فى نهاية عام ٢٠٠٤ - تكشيف ما يقرب من ٣ بليون صفحة (SULLIVAN, 2005, Notess, 2005). ونظراً لأن محركات البحث تعتمد فى كثير من الأحيان على الروابط الفائقة للتعرف على الصفحات والمواقع الجديدة، فقد أوضحت سوزان فيلدمان (Feldman, 1999) أنه من الصعب أن يتم تكشيف صفحة ويب لا تتضمن أى روابط فائقة، كما أوضحت فى دراستها أن محركات البحث تستغرق فى المتوسط ٥٧ يوم لكى تتعرف على أى صفحة جديدة تضاف إلى الشبكة العنكبوتية.

من العرض السابق يتضح أن موضوع التغطية يعد واحداً من أهم المعايير التى يعتمد عليها عند تقييم أداء محركات البحث. هذا على الرغم من أن

(١٠) لا يوجد فرق جوهري بين الأدلة والصفحات الصفراء وإن كانت الأدلة أكثر شمولاً وتعقيداً.

والتكرار في تغطية هذه المجالات بين الأدلة العربية التي سوف نعتمد عليها في تجميع قائمة عشوائية بصفحات المعلومات العربية، بالإضافة إلى معدلات التداخل والتكرار بين المحركات والأدلة محل البحث بهدف تحديد المحركات البؤرية أو الأساسية التي يمكن الاعتماد عليها من جانب الباحث العربي.

#### ٤ - مشكلة الدراسة :

تزايد عدد المستفيدين في العالم بشكل عام وفي الوطن العربي بشكل خاص من الشبكة العنكبوتية كمصدر سهل وسريع للحصول على المعلومات، في الوقت ذاته هناك تزايد مستمر في عدد أدوات البحث المتاحة على الشبكة لتيسير سبل الوصول إلى صفحات ومواقع الويب. يشير برادلي (Bradly, 2000) أنه بحلول عام ٢٠٠٠ كان هناك أكثر من ٢٠٠٠ أداة بحث بعضها عام في تغطيتها والبعض الآخر متخصص في مجالات بعينها. ومع ظهور نظام الشفرة الموحدة (UNICODE) التي تيسر التعامل مع أكثر من ٥٠٠ لغة من اللغات المتاحة في العالم<sup>(١١)</sup>، أصبح من السهل على محركات وأدلة البحث التي تستخدم نظام الشفرة الموحدة في أنظمة تشغيل حاسبتها المركزية أن تقوم بتجميع وتكشيف واسترجاع صفحات ومواقع الويب المتاحة باللغة العربية (UNICODE, INC, 2004). بيد أن تعدد تلك الأدوات وتنوع اهتماماتها أدى إلى صعوبة تعرف المستخدم العربي على أفضلها من حيث شمول التغطية، فالدراسات والقياسات البيولوجرافية

محدودة (إن لم تكن معدومة في هذه الناحية). هذا إلى جانب أن كل أدوات البحث العربية المتاحة على الشبكة العنكبوتية ما زالت بدائية وضعيفة إلى حد كبير من حيث شمول تغطيتها لصفحات ومواقع البيانات العربية أو من حيث إمكانات البحث التي توفرها. فمعظم هذه الأدوات لا تعدو عن كونها أدلة بحث محدودة في نطاق تغطيتها أو أدلة بحث مدعمة بإمكانيات البحث في واحد من محركات البحث العالمية التي توفر إمكانات تقوية إمكانات البحث (Power Your Search By) (داليا نصار، ٢٠٠٤).

من ثم فإن مشكلة هذه الدراسة تتمثل في سعيها نحو الإجابة على السؤال الأساسي المتعلق بتغطية صفحات الويب العربية، بهدف التعرف على أكثر أدوات البحث تغطية لصفحات المعلومات العربية ومدى تفرد هذه الأدوات في تغطية أحد المجالات موضوع الدراسة، وذلك بهدف تقييم تلك الأدوات لمعرفة أكثرها فعالية في البحث عن مصادر المعلومات العربية ووسائل رفع معدلات الاستدعاء Recall في النتائج المسترجعة من أدوات بحث الشبكة العنكبوتية.

#### ٥ - أهمية الدراسة:

تستمد هذه الدراسة أهميتها من كون أدلة ومحركات البحث هي الأدوات الأساسية للوصول إلى مصادر المعلومات المتاحة على الشبكة العنكبوتية. بالإضافة إلى أنها أكثر الأدوات قدرة

(١١) يبلغ عدد اللغات في العالم حوالى ٦٠٠٠ لغة بعضها مكتوب والبعض الآخر غير مكتوب.

i. ما معدلات التداخل والتكرار بين أدلة البحث العربية التي تم الاعتماد عليها في تجميع صفحات المعلومات العربية؟

ii. هل هناك فروق إحصائية في تغطية صفحات المعلومات العربية بين محركات وأدلة البحث التي تشملها الدراسة؟

iii. هل هناك فروق إحصائية في تغطية أدوات البحث التي تشملها الدراسة لصفحات المعلومات العربية في ثلاث مجالات موضوعية هي الإسلاميات، الحاسب الآلي، الصحة؟

#### ٧ - فروض الدراسة:

تسعى هذه الدراسة إلى قياس الفروق الإحصائية بين العناصر الموضحة في تساؤلات الدراسة، لتحقيق ذلك تم صياغة ثلاثة فروض يمكن قياسها إحصائياً وهذه الفروض هي:

١ - يوجد قدر كبير من التداخل والتكرار بين أدلة البحث العربية التي تقوم بتجميع وتنظيم صفحات المعلومات العربية في الموضوعات محل البحث.

للإسلاميات = للحواسيب = للصحة

للإسلاميات = للحواسيب = للصحة

٢ - تنزع محركات البحث العالمية إلى أن تكون

على تجميع وتحليل وتكثيف أكبر قدر ممكن من صفحات المعلومات المتاحة على الشبكة العنكبوتية، سواء كانت باللغة العربية أو باللغات الأخرى. على الرغم من الاختلاف الواضح بين أدلة ومحركات البحث في طرق التجميع والتكثيف والبحث عن صفحات المعلومات إلا أن كلا منهما يلعب دوراً مكتملاً للآخر في تيسير سبل الوصول إلى صفحات ومواقع الويب. وأكبر دليل على ذلك أن الكثير من أدلة البحث أصبحت الآن تدعم قدرتها البحثية من خلال الاعتماد على أحد محركات البحث الشهيرة، كما أن معظم محركات البحث العالمية أصبحت تشتمل الآن على أدلة بحث كوسيلة ثانوية للوصول إلى صفحات ومواقع الويب<sup>(١٢)</sup>. من ثم فإن محركات وأدلة البحث هي الأدوات الأساسية في تيسير عمليات بحث واسترجاع من الشبكة العنكبوتية، وبالتالي فالوقوف على مدى تغطية صفحات المعلومات العربية في تلك الأدوات يساعد بالدرجة الأولى في الإجابة على سؤال هام وملح من جانب مستخدمي تلك الأدوات وهو ما أكثر الأدوات شمولاً في تكثيف وتيسير سبل بحث صفحات المعلومات العربية؟

#### ٦ - تساؤلات الدراسة:

يتمثل الهدف الرئيسي لهذه الدراسة في تقييم التغطية في مجموعة من أدوات البحث العالمية وذلك من خلال الإجابة عن التساؤلات الآتية:

(١٢) انظر دليل البحث جوجل <http://www.google.com/dirhp?hi=ar&tap=wd&q=> وهو عبارة عن دليل بحث يدعم إمكانات التصفح في محرك البحث جوجل. أما الموقع <http://dir.yahoo.com> وهو عبارة عن ذلك بحث Yahoo مدعم بإمكانات البحث الخاصة بمحرك البحث <http://alltheweb.com>.

تغطية قواعد البيانات البليوجرافية في بداية السبعينات والثمانينات من القرن الماضي، وما زال هو المنهج الأساسي في كل دراسات التغطية وقياسات معدلات النمو وحجم محركات وأدلة البحث المتاحة على الشبكة العنكبوتية. حيث تم الاعتماد عليه في العديد من الدراسات المشابهة لهذه الدراسة ومنها ما يلي:

(Bharat & Broder, 1998; Lawrence & Giles, 2002; Vaughan & Thelwall, 2004; Sullivan, 2005 a;)

#### ★ مجتمع الدراسة:

لاختبار الفروض السابقة تم تحديد مجتمع الدراسة في الموضوعات التالية:

- أدلة البحث العربية التي تم الاعتماد عليها لتجميع عينة عشوائية من صفحات المعلومات العربية.
  - محركات وأدلة البحث العالمية التي تم اختبار تغطيتها للصفحات العربية التي تم تجميعها من أدلة البحث السابقة.
  - المجالات المعرفية التي تم اختبار مدى تغطيتها في محركات وأدلة البحث العالمية.
- وسنعرض فيما يلي لأسباب اختبار كل موضوع من هذه الموضوعات.

#### ★ أدلة البحث العربية:

تم اختيار 7 أدلة بحث عربية لتكون الأداة الأساسية لتجميع صفحات المعلومات العربية هي:

أكثر شمولاً في تغطيتها لمواقع المعلومات العربية من أدلة البحث العالمية.

μ محركات البحث العالمية = μ أدلة البحث العالمية

μ محركات البحث العالمية # μ أدلة البحث العالمية

٣ - يوجد قدر كبير من التفرد في تغطية محركات وأدلة البحث العالمية لمجالات الموضوعية التي تشملها الدراسة.

μ إسلاميات = μ حاسب آلي = μ صحة

μ إسلاميات # μ حاسب آلي # μ صحة

#### ٨ - المنهج والإجراءات:

##### ★ المنهج:

قامت هذه الدراسة على تطبيق المنهج التجريبي الذي يعتمد على ضبط جميع متغيرات الدراسة والتحكم فيها للتحقق من صحة فروض الدراسة. وتوضح ملامح المنهج التجريبي في هذه الدراسة في ضبط الثوابت Constants المتمثلة في أدوات البحث التي يتم قياس مدى تغطيتها، المجالات الموضوعية، قائمة ثابتة بصفحات الويب التي تم تجميعها لقياس معدلات تغطيتها، وقياس متغير Variable واحد فقط لكل فرض من فروض الدراسة. وتجدر الإشارة إلى أن المنهج التجريبي هو المنهج الذي تم استخدامه في دراسات التغطية سواء المتعلقة بالأدوات المطبوعة مثل المقارنات التي تمت بين خدمات التكشيف والاستخلاص في بداية الستينات، كما تم الاعتماد عليه في المقارنة بين

١- فارس نت<sup>(١٣)</sup> <http://eyoon.com/206/193/>

٢- الردادى <http://www.raddadi.com/>

٣- عرب لنكس<sup>(١٤)</sup> [www.arablinks.wa](http://www.arablinks.wa)

٤- نسناس <http://www.nesnas.com/>

٥- دليل نسيج <http://directory.naseej.com/>

٦- بوابة العرب <http://www.arabsgate.com/>

٧- شامل نت <http://www.shamel.com/>

وقد قام الباحثان باختيار هذه الأدلة من خلال تجربة رائدة شملت مراجعة القوائم التي تحصر محركات وأدلة البحث العربية. وقد اتضح من المراجعة الدقيقة لهذه القوائم أنه يوجد ٧٠ دليل بحث يهتم بتجميع الصفحات والمواقع العربية. وقد عمد الباحثان إلى اختيار ١٠٪ من هذه الأدلة بحيث تم اختيار أكبر الأدلة العربية وأشهرها وذلك بالاعتماد على أسلوب العينة القصدية أو الغرضية. حيث سعى الباحثان من خلال هذا الاختيار إلى تجميع أكبر قدر ممكن من الصفحات والمواقع العربية بالتالي كان لا بد من اختيار أكثر الأدلة شمولاً. وقد تمت هذه المرحلة بعد مراجعة شاملة للأدلة المتاحة في قوائم أدوات البحث العربية<sup>(١٥)</sup>. وقد ثبت من هذه المراجعة أن الأدلة المختارة هي

(١٣) تغير اسمه إلى عيون مع بداية عام ٢٠٠٥.

(١٤) اندمج مع الدليل فارس نت (عيون) مع بداية عام ٢٠٠٥.

أكبر وأشمل أدلة البحث العربية سواء من حيث الشمول الموضوعي (عدد رؤوس الموضوعات الرئيسية والثانوية المستخدمة في عمليات التكشيف) أو الشمول في عدد الصفحات المكشوفة.

### ★ محركات وأدلة البحث العالمية:

تم اختيار أربع أدوات بحث عالمية لمقارنة تغطيتها لصفحات المعلومات العربية التي تم تجميعها في مجالات الإسلاميات والحاسب الآلي والصحة من أدلة البحث العربية. وقد اشتملت هذه الأدوات على محركين عالميين هما Google, Alta Vista ودليلين من أدلة البحث الشهيرة أيضاً على المستوى العالمى هما Yahoo, Look Smart.

حيث تشير الإحصائيات إلى أن Google هو أكثر المحركات تغطية لصفحات المعلومات على المستوى العالمى وأن قاعدة بياناته هي أكبر قاعدة بيانات ضمن قائمة المحركات العالمية، حيث تتضمن ما يقرب من ١٠ بليون تسجيله بينما يأتي Alta Vista فى الترتيب الثانى من حيث الشهرة العالمية والأصالة، حيث أنه من أقدم محركات البحث على المستوى العالمى بينما يأتي فى الترتيب الثالث من حيث حجم قاعدة بياناته وعادة ما يدخل فى منافسة شديدة مع محرك البحث

(١٥) قائمة بأشهر محركات وأدلة البحث العربية <http://aslimnet.free.fr/services/moteurs.htm> وتشتمل هذه القائمة على ٧٥ أداة بحث تام استبعاد ٢٠ منهم نظراً لأنهم أدوات متحيزة جغرافية أو موضوعية مثل دليل المواقع المصرية، دليل المرأة الخليجية. والقائمة الثانية هي محركات البحث العربية والعالمية <http://www.geocities.com/qurna1/search1.html> وتشتمل على ٤٥ محرك ودليل بحث عالمى وعربى. وبمراجع القائمتان القائمتان واستبعاد المكررات تم التركيز على ٧٠ دليل فقط تمثل النواة الأساسية لاختيار ٧ أدلة عامة.

لعدد المواقع في مختلف المجالات المعرفية حيث اتضح ما يلي:

أن هذه المجالات الثلاثة موضوعاً أساسياً في أي دليل بحث عربي مما يسر عملية تجميع الصفحات. كما أن عدد صفحات المعلومات المتاحة تحت هذه المجالات كاف بشكل كبير لإجراء مثل هذه الدراسات التي تحتاج إلى عدد كبير من الصفحات. وقد تم الكشف عن هذا من خلال تجربة رائدة Pilot Experiment أجريت لتحليل أدلة البحث العربية المتاحة على شبكة الإنترنت حيث أتضح أن هذه المجالات هي أكثر المجالات التي لها صفحات معلومات باللغة العربية. فقد أوضحت هذه التجربة أن عدد الصفحات المتاحة في هذه المجالات الثلاثة يقرب من ٢٠ ألف صفحة (١٩٧٩٠ صفحة) في كل الأدلة محل الدراسة.

#### ★ اختيار العينة:

تم اختيار عينة عشوائية لصفحات الويب المتاحة باللغة العربية وذلك على مرحلتين أساسيتين كما يلي:

#### \* المرحلة الأولى:

التعرف على أدلة البحث التي تغطي مواقع الويب العربية وحصرها حصراً دقيقاً لاختيار أكثر الأدلة العربية تغطية للصفحات والمواقع العربية<sup>(١٦)</sup>.

#### \* المرحلة الثانية:

الاختيار العشوائي لصفحات الويب المتاحة من خلال أدلة البحث باستخدام جداول الاختيار

Alltheweb (www.alltheweb.com) المركز الثاني ضمن قائمة محركات البحث العالمية (Sullivan, D, 2005a). ومن المعروف والثابت في كثير من الدراسات والمواقع التي تهتم بأدوات البحث على الشبكة العنكبوتية أن كلا من Yahoo, Look Smart من أشهر الأدلة وأكثرها ثباتاً من حيث الأداء، كما أنهما يتمتعان بإمكانيات وقدرات بحث تضاهي إلى حد كبير إمكانيات البحث التي توفرها محركات البحث العالمية. ويعتبر دليل البحث Yahoo من أقدم أدلة البحث وقد أُضيف إلى الدليل إمكانيات البحث من خلال محرك البحث Alltheweb (Sullivan, D.) (2005b, July 2). وهذا هو السبب الرئيسي وراء استبعاد محرك البحث Alltheweb من الدراسة لأنه يدعم إمكانيات البحث في Yahoo. كما أضاف دليل البحث Look Smart إمكانيات البحث في الدليل من خلال المحرك Alltheweb أيضاً. وتجدر الإشارة إلى أن Yahoo بدء كدليل بحث ولكنه تحول من محرك بحث يتضمن كل مكونات محرك البحث، إلا أن الدراسة اعتمدت في تقييمه على البحث داخل الدليل وليس البحث باستخدام المحرك.

#### ★ المجالات المعرفية:

تم تجميع عينة عشوائية من صفحات المعلومات العربية اشتملت على ١٨٧٥ صفحة موزعة على ثلاث مجالات معرفية هي الإسلاميات، الحاسب الآلي والإنترنت، الصحة. وخضع اختيار هذه المجالات الثلاثة إلى دراسة رائدة Pilot Study

(١٦) تجدر الإشارة أن الدراسة تعاملت فقط مع الصفحات بمفهومها المستقل ولم تتطرق إلى المواقع بمفهومها التجميعي.

العشوائية حيث تم تحديد الصفحات المختارة عشوائياً واستبعاد المواقع التي ما زالت تحت الإنشاء والمواقع التي تبين الأسماء السائدة Domain Names والمواقع التي تحتاج إلى كلمات مرور من أجل تصفحها حيث أنها غير متاحة لمحركات البحث والمواقع التي تستخدم الكود Txt Robot وهو عبارة عن أحد أكواد لغة تكويد النصوص

الفائقة HTML الذي لا يسمح لمحركات البحث باصطياد الصفحة وتكشيفها، هذا بالإضافة إلى استبعاد أى مواقع تم إنشائها باللغات غير العربية.

وقد تم تجميع ١٨٧٥ صفحة معلومات - خلال شهرى سبتمبر وأكتوبر من عام ٢٠٠٤ - موزعة على ثلاثة مجالات معرفية يوضحها الجدول التالي:

النسبة المئوية	عدد الصفحات التي تم تجميعها	العدد المطلوب	عدد الصفحات	المجال الموضوعى
٧٤٢	٧٩٠	٨٣٢	٨٣١٢	الإسلاميات
٧٣٦	٦٧٠	٧١٢	٧١٢٤	الحاسب والإنترنت
٧٢٢	٤١٥	٤٣٥	٤٣٥٤	الصحة
٧١٠٠	١٨٧٥	١٩٧٩	١٩٧٩٠	المجموع

يوضح الجدول السابق أنه تم تجميع ٧١٠ تقريباً من الصفحات المتاحة فى المجالات المعرفية الثلاثة والذي يبلغ ١٩٧٩. ويرجع التفاوت فى عدد الصفحات المجمع فى المجالات الثلاثة إلى التفاوت فى عدد الصفحات المتاحة لكل مجال من المجالات الثلاثة حيث أتى مجال الإسلاميات فى المرتبة الأولى من حيث عدد الصفحات، يليه مجال الحاسب الآلى والإنترنت ثم أخيراً مجال الصحة كما هو واضح فى الجدول السابق.

تم تحديد حجم العينة بالاعتماد على برنامج متاح على شبكة الإنترنت لتقدير حجم العينات بعد تقدير حجم المجتمع الكلى للمواقع المتخصصة فى المجالات محل الدراسة حيث تم إدخال الحجم الكلى للمجتمع ومستوى الثقة المطلوب Level of Confidence فى نتائج الدراسة، وتم الحصول على تقدير لحجم العينة Sample Size Determination كما هو موضح بالجدول التالى الذى تم نسخه من برنامج تقدير حجم العينة لمجتمع معروف ومحدد.

## Sample Size with Acceptable Absolut Precision for Finite Populations

Pilot Sample Size (n):	45
Poulations Size (N):	19790
Current Variance Estimate:	0.5
Acceptable Significant Level ( $\alpha$ ):	.05
Acceptable Absolut Error:	.03
The Required Sample Size Is:	1979

جدول رقم (٢) تحديد حجم العينة باستخدام برنامج تقدير حجم العينات  
في المجتمعات المحددة

الدراسة كما هو موضح في جدول رقم (١).

### \* خطوات الدراسة:

لتحقيق الغرض من الدراسة قام الباحثان باتباع مجموعة الخطوات التالية:

i. مسح لأدلة البحث العربية المتاحة على شبكة الإنترنت وتحديد أكثرها شمولاً في تغطية مصادر وصفحات المعلومات العربية، لتكون نواة يمكن الاعتماد عليها لتجميع قائمة بصفحات المعلومات العربية المتاحة على الشبكة في الثلاث مجالات محل الدراسة.

ii. مسح كل دليل من الأدلة المختارة لتجميع قائمة شاملة بمصادر المعلومات العربية التي يشملها هذا الدليل في كل مجال من المجالات المذكورة على حدة.

وكما هو واضح من الجدول أن الحجم الإجمالي للمجتمع بلغ ١٩٧٩٠ صفحة. وقد تم تقدير هذا الحجم بناء على الدراسة الميدانية من خلال تقدير متوسط عدد الصفحات في كل مجال موضوعي يغطيه الدليل ثم متوسط عدد الرؤوس الثانوية ثم متوسط عدد الصفحات في كل صفحة بها رأس ثانوي. ويوضح الجدول أن حجم عينة الدراسة يبلغ ١٩٧٩ صفحة بمستوى ثقة ٠,٩٥. وقد تم تجميع ٢٠٠٠ صفحة في ثلاث ملفات موزعة على مجالات الدراسة وبعد فحص المواد المجمعة تم استبعاد ١٢٥ صفحة نظراً لأنها صفحات لها روابط فائقة في أدلة البحث ولكنها لم تعد متاحة على الشبكة العنكبوتية فيما يعرف بالروابط الميتة Dead Links كما تم استبعاد الصفحات التي تتضمن الكود Robot.txt والمواد بغير اللغة العربية. بالتالي بلغ الحجم الإجمالي لعدد الصفحات التي تم فحصها ١٨٧٥ صفحة موزعة على مجالات



iii. دمج نتائج التجميع في ثلاث ملفات مستقلة حيث يمثل كل ملف مجال موضوعي من المجالات الثلاثة.

iv. تشغيل برنامجين تم إعدادهما خصيصا لهذه الدراسة باستخدام لغة البرمجة Perl لإعداد ملف كشاف Index HTML File يشتمل على قائمة شاملة بالصفحات الفريدة التي تم تحديد مدى تغطيتها في أدلة ومحركات البحث العالمية. يعمل البرنامج الأول على استبعاد الصفحات المكررة والاحتفاظ بنسخة واحدة من كل صفحة في الملف الكشاف. وذلك من خلال المقارنة بين معين المصادر الموحد Uniform Resource Locator (URL) للصفحات، فإذا تكرر العنوان يتم استبعاد المكررات والاحتفاظ بنسخة وحيدة في الملف. ويعمل البرنامج الثاني على تحديد أرقام الصفحات المكررة بين الأدلة العربية بالتالي يمكن قياس معدلات التداخل والتكرار بين الأدلة العربية.

v. استخدام البيانات المتاحة في كل ملف من الملفات الكشافية في قياس التغطية. وقد تم إعداد ثلاث ملفات بحيث أشتمل كل ملف على الصفحات الفريدة - بعد استبعاد المكررات وقياسها من خلال البرامج المذكورة أعلاه - الخاصة بكل مجال من مجالات الدراسة

لكي يتم الاعتماد عليها في قياس التغطية في أدوات البحث العالمية التي تضمنتها الدراسة.

vi. إعداد قائمة بكل الصفحات التي تم اكتشافها في كل أداة بحث على حدة للتعرف على مدى الشمول والتفرد في تغطية المجالات محل الدراسة وذلك من خلال البحث عن الصفحات التي تم تجميعها في أدوات البحث محل الدراسة باستخدام معين المصادر الموحد كعامل البحث Search Operator.

vii. استخدام نظام درجات Scoring System يحدد موقع ظهور الصفحة التي يتم البحث عنها عند إجراء البحث في الأدوات محل الدراسة (سواء كانت الأدلة أو محركات البحث). واشتمل هذا النظام على أربع درجات منفصلة Discrete Score أو غير خطية Non Linear Score. وهذه الدرجات هي:

- ٣ ← إذا ظهرت الصفحة التي يتم البحث عنها ضمن الترتيب من ١ - ٣
- ٢ ← إذا ظهرت الصفحة التي يتم البحث عنها ضمن الترتيب من ٤ - ٧
- ١ ← إذا ظهرت الصفحة التي يتم البحث عنها ضمن الترتيب من ٨ - ١٠
- ٠ (صفر) ← إذا لم تظهر الصفحة التي يتم البحث عنها ضمن العشر نتائج الأولى.

## ٩ - مناقشة النتائج

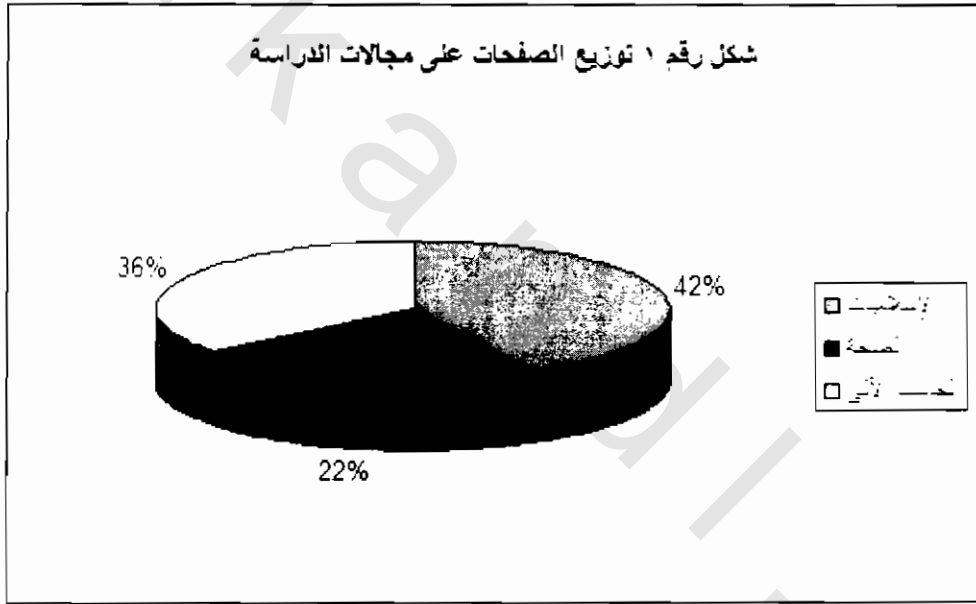
### :Discussion of Results

فيما يلي عرض لأهم النتائج التي توصلت إليها الدراسة مرتبة وفقاً لارتباطها بأسئلة وفروض الدراسة:

### ★ أولاً توزيع الصفحات والتداخل والتكرار بين الأدلة العربية:

تضمنت الدراسة عدد ١٨٧٥ صفحة

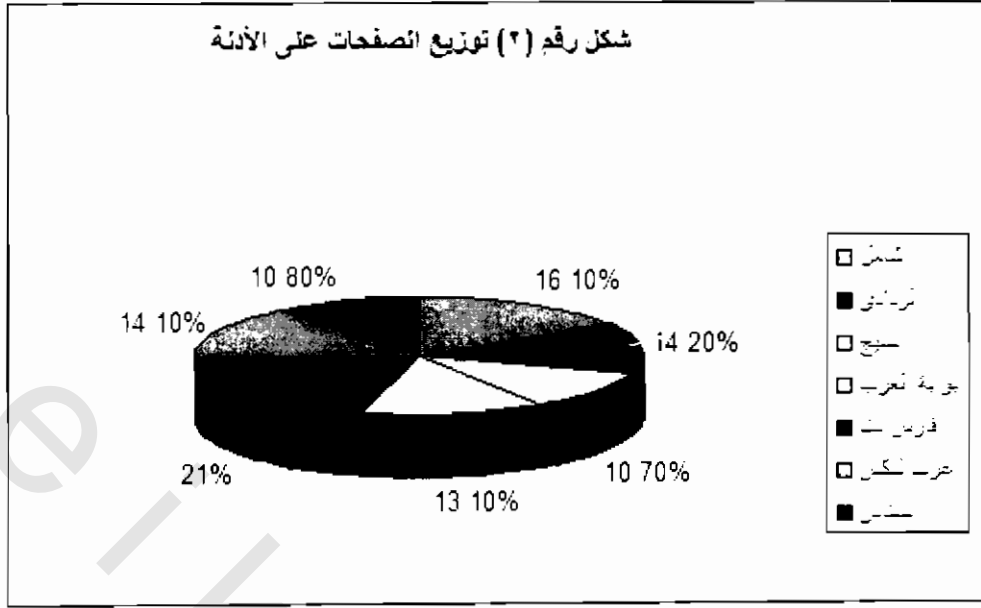
معلومات عربية تم تجميعها من ٧ أدلة بحث عربية. وقد تفاوتت نسب تمثيل تلك الصفحات للمجالات الموضوعية الثلاثة محل الدراسة حيث تضمنت الدراسة ٤١٥ صفحة في مجال الصحة بنسبة ٢٢٪، و ٦٧٠ صفحة في مجال الحاسب الآلي بنسبة ٣٦٪، و ٧٩٠ صفحة في مجال الإسلاميات بنسبة ٤٢٪ من إجمالي الصفحات محل الدراسة. كما هو موضح في الشكل (١).



المعلومات في مجال الحاسب الآلي والإنترنت ثم مصادر المعلومات المرتبطة بمجال الصحة. وتجدر الإشارة إلى أنه تم استبعاد الكثير من المصادر المتعلقة بمجال الصحة نظراً لكونها معدة باللغة الإنجليزية مما أدى إلى تقلص نسبة تمثيل تلك الفئة الموضوعية، إلا أن إضافة تلك الصفحات لا يؤثر كثيراً في تفوق مجالات الإسلاميات وعلوم الحاسب الآلي.

ويرجع هذا التفاوت إلى اختلاف نسب التغطية لصفحات المعلومات من قبل أدلة البحث العربية في تلك المجالات الموضوعية. فعلى الرغم من تحديد نسبة مئوية موحدة (٢٠٪) من عدد الصفحات المتوافرة بكل دليل) للحصول على صفحات المعلومات من جميع أدلة البحث المختارة بالدراسة (٧ أدلة) إلا أنه قد ظهر بوضوح انحياز تلك الأدلة لتغطية مصادر المعلومات الإسلامية يليها مصادر

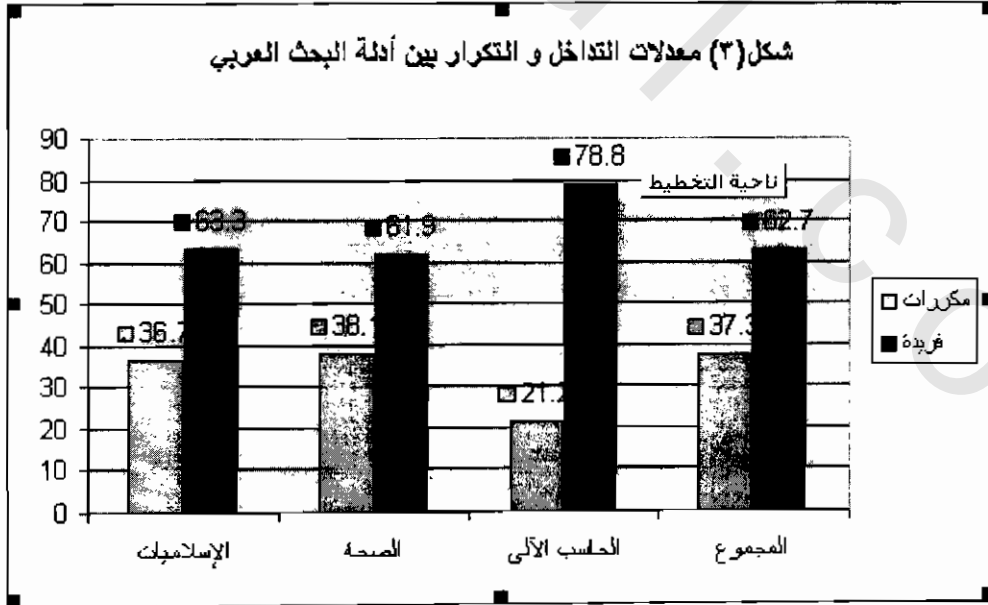
شكل رقم (٢) توزيع النصفحات على الأدلة



أنه أقدم أدلة البحث العربية حيث يعتبر من المشروعات الرائدة لبناء أدلة بحث عربية بالإضافة إلى أنه أكثر الأدلة نمواً في السنوات الأخيرة حيث قام بضم دليل البحث عرب لنكس مما ساعد على نمو دليل البحث فارس نت (عيون).

ويوضح الشكل (٢) أن دليل البحث فارس نت هو أكثر الأدلة شمولاً في تغطية صفحات المعلومات العربية يليه دليل البحث شامل نت ثم دليل البحث الرادى وعرب لنكس، بينما تتقارب باقي الأدلة الأخرى في تغطية وتكثيف صفحات المعلومات العربية. ويرجع تفوق دليل فارس نت إلى

شكل (٣) معدلات التداخل و التكرار بين أدلة البحث العربي



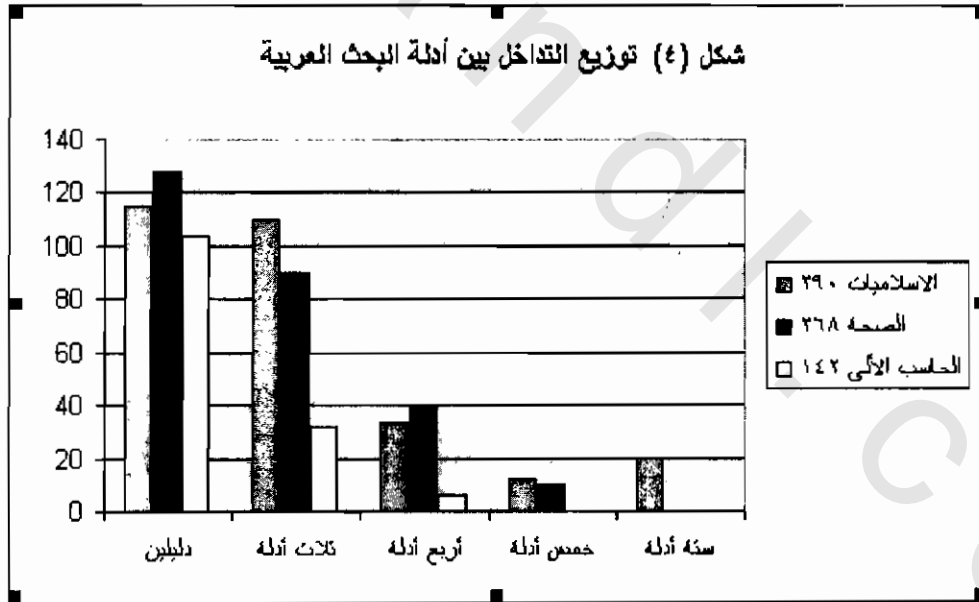
مقارنتها بعدد الصفحات على المستوى العالمي. كما يرجع أيضاً ارتفاع معدلات التداخل والتكرار في مجال الصحة إلى أنه أقل المجالات التي لها صفحات على الويب عند مقارنتها بمجالات الإسلاميات والحاسب الآلي.

كما أوضحت الدراسة أن أعلى معدلات التداخل والتكرار بين الأدلة تمثلت في مجال الصحة عند قياس التداخل بين دليلين وأربع أدلة فقط بينما تمثلت في مجال الإسلاميات عند قياس التداخل بين ثلاث وخمس وست أدلة، بينما كان التداخل في مجال الحاسب الآلي أقل بكثير من المجالين الآخرين كما هو موضح في الشكل (٤).

ويوضح الشكل (٣) أن معدل التداخل والتكرار بين أدلة البحث بصفة عامة يبلغ ٣٧,٣٪ بينما ٦٢,٧٪ من الصفحات هي صفحات فريدة وغير مكررة في أي من أدلة البحث. ويبلغ أعلى معدل للتكرار بين الأدلة في مجال الصحة ٣٨,١٪ يليه مجال الإسلاميات ٣٦,٧٪ وأخيراً مجال الحاسب الآلي ٢١,٢٪. وترجع الزيادة الكبيرة في معدلات التداخل والتكرار بين أدلة البحث العربية - التي تفوق معدلات التداخل والتكرار بين أدوات البحث العالمية (انظر

<http://www.searchengineshowdown.com/stats/overlap.shtml>

إلى أن حجم الصفحات العربية أقل بكثير عند



غير مرتبطة بحجم التغطية في أدلة البحث العربية. والدليل على أن مجال الإسلاميات هو أكثر المجالات تغطية في كل أدلة البحث العربية إلا أن معدلات

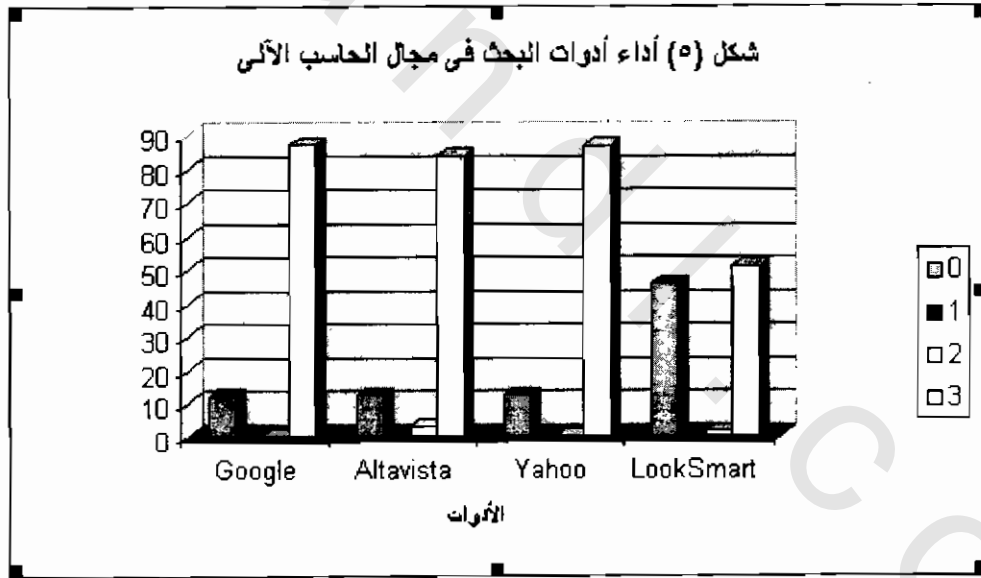
ويتضح من هذه التوزيعات صحة الفرض الأول الذي وضعته الدراسة وهو أن معدلات التداخل والتكرار تختلف من مجال لآخر بالإضافة إلى أنها

العالمية للمصادر العربية اقتصرَت الدراسة على النتائج العشرة الأولى لعملية البحث لكل مصدر من المصادر المختارة وعددها ١٨٧٥ صفحة. وقد تم استخدام نظام درجات Scoring System يحدد القيمة وفقاً لموقع الصفحة التي يتم البحث عنها ضمن النتائج العشر الأولى كما يلي: القيمة صفر في حالة عدم ظهور الصفحة في النتائج العشرة الأولى، القيمة ١ في حالة ظهور الصفحة ما بين ٨ - ١٠ في ناتج عملية البحث، القيمة ٢ في حالة ظهور الصفحة ما بين ٤ - ٧، والقيمة ٣ في حالة ظهور الصفحة ما بين ١ - ٣ من نتائج عملية البحث. وقد أسفرت نتائج الدراسة عما يلي:

التداخل والتكرار بين الأدلة في هذا المجال أقل من مجال الصحة على الرغم من أن مجال الصحة أقل المجالات تغطية في أدلة البحث العربية. كما أن معدلات التداخل في مجال الحاسب الآلي أقل من معدلات التداخل في مجال الصحة على الرغم من أن عدد الصفحات في مجال الحاسب والإنترنت أكبر بكثير من عدد الصفحات في مجال الصحة. هذا وإن كان مجال الإسلاميات هو المجال الوحيد الذي ظهر في كل نماذج التداخل (من دليلين إلى ستة أدلة).

### ★ ثانياً فروق تغطية محركات وأدلة البحث للمصادر العربية:

للتعرف على تغطية محركات وأدلة البحث



توضح موضع الصفحة ضمن مجموعة الصفحات المسترجعة، يليه دليل البحث Yahoo، ثم محرك البحث Alta Vista وأخيراً دليل البحث Look Smart (كما هو موضح بالشكل السابق).

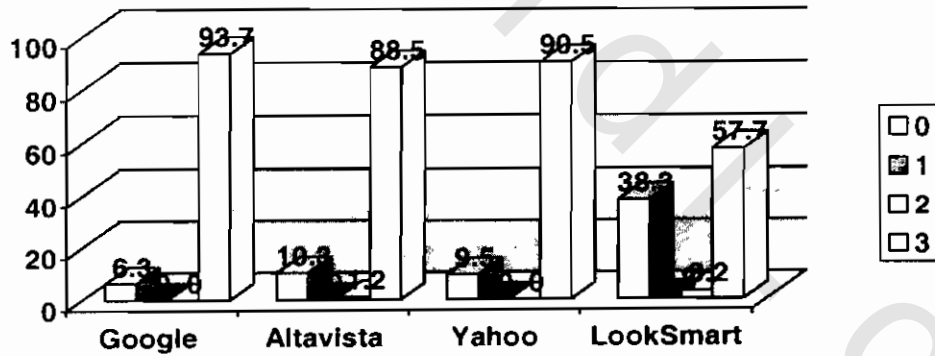
محرك البحث Google هو أكثر المحركات تغطية لصفحات المعلومات العربية في مجال الحاسب الآلي وذلك للصفحات التي تظهر ضمن النتائج من ٣-١ أي التي حصلت على القيمة ثلاثة التي

جدول رقم (٣) تغطية المصادر العربية في مجال الحاسب الآلى

الانحراف المعياري	الوسط الحسابي	المحرك / الدليل
,٩٩	٢,٦٢	Google
١,٠٠	٢,٦١	Yahoo
١,٠٠	٢,٥٨	Alta Vista
١,٤٨	١,٥٧	Look Smart

وكما هو الحال في مجال الحاسب والإنترنت أوضحت الدراسة أن محرك البحث Google هو أكثر أدوات البحث تغطية من حيث ظهور المصادر العربية ضمن النتائج من ١ - ٣ لعملية البحث في مجال الصحة، يليه دليل Yahoo، ثم Alta Vista، وأخيراً Look Smart.

شكل رقم (٦) تغطية محركات وأدلة البحث في الصحة



المتوسطات الحسابية والانحرافات المعيارية كما هو موضح بالجدول التالي:

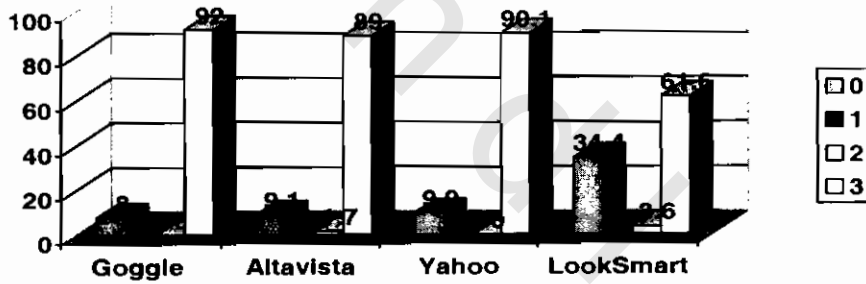
وقد تشابهت النتائج في مجال الصحة بشكل عام مع مثيلتها في مجال الحاسب الآلى فيما يتعلق بترتيب محركات وأدلة البحث مع اختلاف

جدول رقم (٤) تغطية المصادر العربية في مجال الصحة

الانحراف المعياري	الوسط الحسابي	المحرك / الدليل
,٧٣	٢,٨١	Google
,٨٨	٢,٧٢	Yahoo
,٩٢	٢,٦٨	Alta Vistal
١,٤٤	١,٨٠	Look Smart

وفي مجال الإسلاميات أثبتت النتائج أيضاً أن محرك البحث Google هو الأكثر تغطية يليه محرك البحث Alta Vista ثم دليل البحث Yahoo وأخيراً دليل Look Smart .

شكل رقم (٧) تغطية محركات وأدلة البحث في الإسلاميات

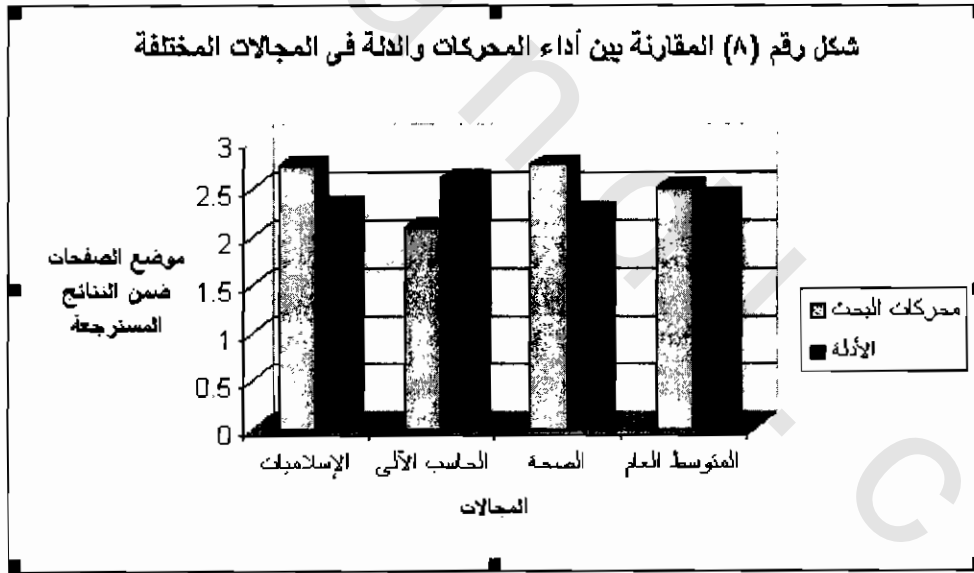


جدول رقم (٥) تغطية المصادر العربية في مجال الإسلاميات

الانحراف المعياري	الوسط الحسابي	المحرك / الدليل
,٨٢	٢,٧٦	Google
,٨٧	٢,٧١	Alta Vistal
,٩٠	٢,٧٠	Yahoo
١,٤١	١,٩٢	Look Smart

أوسع لصفحات المعلومات في مجال الحاسب الآلي كما هو موضح في الشكل التالي.

وقد أثبتت الدراسة أن محركات البحث أكثر تغطية لصفحات المعلومات العربية من أدلة البحث وذلك في المتوسط العام بينما تنفرد الأدلة بتغطية



بالتغطية في كل أدوات البحث محل الدراسة. هذا وإن كانت التحليلات الإحصائية التي اعتمدت على استخدام اختبار t (T Test) أثبتت عدم وجود فروق إحصائية بين محركات وأدلة البحث مما يعني عدم

كما أوضحت النتائج الوصفية العامة أن محرك البحث Google أفضل أدوات البحث تغطية لمصادر المعلومات العربية المتاحة على الشبكة العنكبوتية، وأن مجال الإسلاميات هو أكثر المجالات التي تخطى



صحة الافتراض الثاني والذي يفترض أن محركات البحث تنزع لأن تكون أفضل من أدلة البحث في تغطية صفحات المعلومات العربية كما هو واضح في الجدول التالي:

جدول رقم (٦) تغطية المصادر العربية في مجال الإسلاميات

t-test for Equality of Means						Levene's Test for Equality of Variances			
Confidence 95% Interval of the Difference		Std. Error Difference	Mean Difference	Sig. (2-tailed)	d'	t	Sig.	f	
Upper	Lower								
-0.40707								Equal variances assumed	VAR 00004
-0.53199	0.03186	-0.4695	0	4994	-14.738	0	932.639		
-0.40708	-0.053198	0.03185	-0.4695	0	4461.675	-14.74			Equal variances not assumed

الآلى Automatic Crawling من خلال محرك البحث الذى قام بشرائه. أما تفوق محرك البحث Alta Vista على دليل البحث Look Smart فهو منطقى ومتوقع نظراً لتاريخ إنشاء كلا منهما إلى جانب أن محرك البحث يعتمد على التجميع الآلى بينما دليل البحث يعتمد على التجميع اليدوى.

★ **ثالثاً: التفرد والتميز فى التغطية الموضوعية:**

أثبت اختبار تحليل التباين فى اتجاه واحد (One Way ANOVA) أنه لا يوجد اختلاف بين محركات البحث فى تغطية المجالات الموضوعية محل الدراسة (الإسلاميات، الحاسب الآلى

ويرجع التفوق النسبى لمحرك البحث جوجل عن باقى أدوات البحث والاسترجاع إلى الطريقة التى يستخدمها جوجل فى عمليات التكشيف التى تعتمد على أسلوب تحليل الروابط الفائقة Hyperlinks Analysis والتى تسمح له بتتبع الصفحات المرتبطة بالصفحة التى يكشفها مما يساعد على توسيع نطاق التغطية. ويرجع تفوق دليل البحث Yahoo على محرك البحث Alta Vista إلى جوانب تاريخية تتعلق بتاريخ بدء العمل فى كلا منهما حيث أن Yahoo أقدم نسبياً من Alta Vista بالإضافة إلى اعتماده تكنولوجيا الزحف

والصحة)، حيث ثبت أن قيمة كاي تربيع غير دالة على مستوى أقل من  $\text{Level of Sig.} < 0.5$  ← التالى: التالى: (Sig= .001 < .05) كما هو موضح فى الجدول

جدول رقم (٧) تحليل ANOVA لتغطية محركات البحث

ANOVA

SEARCHEN

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11.087	2	5.543	6.722	.001
Within Groups	2057.626	2495	.825		
Total	2068.713	2497			

من ناحية أخرى أثبت إجراء تحليل التباين فى الاتجاه واحد (One Way AVOVA) أنه لا يوجد اختلاف بين أدلة البحث فى تغطية المجالات الموضوعية محل الدراسة (الإسلاميات، الحاسب الآلى والصحة، حيث ثبت أن قيمة كاي تربيع غير دالة على مستوى أقل من  $\text{Level of Sig.} < 0.5$  ← 0.05) كما هو موضح فى الجدول التالى:

جدول رقم (٨) تحليل ANOVA لتغطية أدلة البحث

ANOVA

DIRRECTO

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	25.170	2	12.585	7.411	.001
Within Groups	4236.650	2495	1.698		
Total	4261.819	2497			

من ثبت عدم صحة الافتراض الثالث الذى يفترض وجود قدر كبير من التفرد فى تغطية محركات وأدلة البحث لمجالات موضوعية، مما يعنى عدم تركيز تلك الأدوات على مجالات بعينها. كما

أثبت تحليل التباين للفرق بين الأدلة ومحركات البحث في تغطية المجالات الموضوعية المختلفة عدم وجود فروق إحصائية بين الأدلة ومحركات البحث في تغطية مجال الإسلاميات والصحة والحاسب الآلى والإنترنت ( $Sig=.001 < .05$ ). هذا على الرغم من تفوق محركات البحث على الأدلة نسبياً ولكن دون وجود فروق إحصائية تميز المحركات عن الأدلة. ويرجع السبب وراء التقارب الكبير بين محركات وأدلة البحث في تغطية صفحات ومواقع الويب إلى السباق المحموم بين تلك الأدوات فى محاولة السيطرة على الشبكة العنكبوتية، كما أن الفروق الكبيرة بين أدلة ومحركات البحث التي ظهرت فى نهاية التسعينات من القرن الماضى بدأت تتلاشى مع محاولة كل أدوات البحث أن تكون بوابات رئيسية للويب مما يجذب أكبر عدد من المستفيدين إليها. حيث أثبت التحليل الدقيق لكلا من جوجل وياهو أنهما بالفعل يتضمننا كل عناصر بوابات الويب Web Portals من خدمات بحث واسترجاع، خدمات بريدية ودردشة، خدمات إخبارية، استفتاءات ومسابقات وغيرها من الخدمات التى تجعل من الموقع البوابة الأساسية للمستفيد على الويب.

### الخاتمة:

أثبتت نتائج الدراسة التفوق النسبى لمحركات البحث على أدلة البحث فى تغطية مصادر المعلومات العربية، بالإضافة إلى انخفاض حجم تمثيل المصادر العربية فى مجال الصحة الذى يعد من أهم المجالات الحيوية للجماهير العام من مستخدمي الشبكة العنكبوتية. وقد خلصت الدراسة إلى وجود قصور

فى كشف أدلة البحث العالمية لصفحات المعلومات العربية مقارنة بمحركات البحث العالمية. وتجدر الإشارة إلى أنه على الرغم من عدم إمكانية تعميم النتائج على جميع مصادر المعلومات العربية المتاحة على شبكة الإنترنت، إلا أن تلك الدراسة يمكن أن تعد بمثابة بداية لعدة دراسات لمصادر المعلومات المتاحة عبر الإنترنت. من ثم توصى الدراسة بما يلى:

- ضرورة اتجاه أدلة البحث العربية نحو التفرد فى تغطية مجالات موضوعية معينة بحيث نتخذ من ظاهرة التكرار وتغطى أكبر قدر ممكن من صفحات المعلومات العربية.
- أن يتجه الباحث العربى إلى استخدام محركات البحث العالمية لتجميع أكبر قدر ممكن من الصفحات العربية التى تلبي احتياجاته المعلوماتية، حيث أوضحت الدراسة أن محركات البحث أكثر شمولاً «نسبياً» لصفحات المعلومات العربية من أدلة البحث وخاصة محرك البحث Google .
- دمج النتائج المسترجعة من كلا من Google و Yahoo أفضل من دمج نتائج كلا من Google و Alta Vista . كما أن دمج نتائج Google, Yahoo, Alta Vista أفضل من دمج نتائج Google, Yahoo, Look Smart بالتالى فإضافة نتائج Look Smart يؤثر سلباً على نتائج الاسترجاع.
- كما توصى الدراسة أيضاً بالتوسع فى كشف المجالات الموضوعية محل الدراسة فى أدلة البحث العربية بالإضافة إلى إنشاء أدلة بحث

استخدامها من قبل العاملين على تلك المواقع لإنتاجها بشكل أفضل ضمن نتائج البحث في تلك الأدوات العالمية.

### المراجع الأجنبية:

1. Bharat, K., & Broder, A. (1998). Measuring the Web (An update to conference paper : Estimating the relative size and overlap of public Web search engines, presented at the 7<sup>th</sup> International World Wide Web Conference in April '98). <http://www.research.compaq.com/src/whatsnew/sem.html> (accessed 5 March 2003).
2. Brier, D.J., et. al. (1999). Evaluating Title Coverage of Full Text Periodicals at the University of Hawaii. The Journal of Academic Librarianship, v. 25, No. 6, pp. 473-478.
3. Bradly, B. (2000). Multi search Engines Comparison. Retrieved from the web at April, 12, 2004 – ERIC Abstract, <http://www.phlib.com/msengine.html>.
4. Chowdhury, G.G. (1999). The Internet and information retrieval research, a brief-review. Journal of documentation, Vol. 55, No. 2, March.

عربية متخصصة فسي تلك المجالات. وأن يتولى القائمون على أدلة البحث العربية عملية تسجيل المواقع - المكشفة بها - بمحركات وأدلة البحث العالمية أو بيع بيانات التكشيف لتلك المحركات العالمية. مع العلم أنه توجد العديد من المواقع العالمية والعربية المتخصصة التي تنمو بشكل كبير وملحوظ وتسعى نحو تطوير البوابات المتخصصة في مجالات موضوعية معينة مثل بوابة إسلام أون لاين <http://www.islamonline.net/english/index.shtml> بوابة الحاسب الآلي [www.thecomputerportal.com/](http://www.thecomputerportal.com/) البوابة الصحية <http://www.thehealthportal.com/>

• ضرورة إجراء دراسات مماثلة لتغطية المصادر العربية في مجالات موضوعية أخرى ومقارنتها بنتائج الدراسة الحالية كما لا بد من إنشاء مراكز أو مؤسسات متخصصة في إجراء دراسات عن سبل دعم التواجد العربي على شبكة الإنترنت الذي يمكن أن يأخذ أشكال متعددة منها أساليب دعم ترتيب الصفحات العربية في محركات البحث العالمية Arabic Web Pages Optimization أساليب بناء المواقع العربية التي تتماشى مع تقنيات البحث في أدوات البحث العالمية وغيرها من المجالات التي تحتاج إلى مراكز للبحث والتطوير.

• ضرورة إجراء دراسة تجريبية لمعرفة أسباب تراجع ظهور بعض المواقع العربية في أدلة ومحركات البحث العالمية. ووضع استراتيجية يمكن

- online at : <http://www.internetworldstats.com/stats.htm>
10. Lancaster, F.W. (1998). Indexing and Abstracting in Theory and Practice. Champaign, Illinois : University of Illinois, Graduate School of Library and Information Science, 412 p.
  11. Lancaser, F.W., Warner, Amy. J. (1993). Information Retrieval Today. Information Resource Press : Arlington, 341 p.
  12. Lawrence, S. & Giles, C.L. (1999). Accessibility of Information on the Web. Nature, Vol. 400, 107-109. Available online (June, 19, 2002) at <http://www.metrics.com/>
  13. Lawrence, S. & Giles C.L. (2002). New Study on the Accessibility and Distribution of Information on the Web. Available Online June, 19, 2002. <http://www.neci.nec.com/~lawrence/searchtips.html>
  14. Lenssen, Phillip. Search Engines History. April, 2004. Retrieved from the WWW at May 14, 2005. <http://blog-outer-court.com/history/>
  15. Line, M.B. An information world apart : The Royal Society Scientific Information conference of 1984 in
  5. Feldman, Susan (1999). New Study of Search Engines Coverage, Information Today, Vol. 16, No. 829. Retrieved from the web at April, 30, 2005. <http://www.infotoday.com/newsbreaks/nb0712-1.htm>
  6. Gromov, Gregory R. (2004). History of Internet and WWW: The Roads and Crossroads of Internet History. Retrieved from the Web at July, 28, 2004. <http://www.nua.ie/surveys/index.html>
  7. Hawking, D., Craswell, N., Bailey, P., Griffith, K. (2001). Measuring Search Engines Quality. Information Retrieval. 4 (1), pp. 33-39.
  8. Henzinger, M.R., Heydon, A., Mitzenmacher, M., & Najork, M. (1999). Measuring index quality using random walks on the Web. In Proceedings of the 8<sup>th</sup> International World Wide Web Conference. [On-line]. Available online at <http://www8.org/w8-papers/2c-search-discover/measuring/measuring.html>
  9. Internet Usage Statistics : Usage and Population Statistics (2005). World Internet Users and Population Statistics. Retrieved from the web at 20, November, 2005. Available

- Showdown. Retrieved August, 2005 From <http://www.search-enginesshowdown.com/stats/size.shtml>
22. Read, E.J. et. al., (2000). Searching for Library and Information Science Literature : A Comparison of Coverage in three Databases (LISA, ISA, and Library Literature & Information Science). Library Computing, V. 19, No. 1, pp. 118-126.
  23. Sullivan, D. (2005a, December 11). Search engine sizes [On-line]. Available : <http://www.searchenginewatch.com/reports/sizes.htm>
  24. Sullivan, D. (2005b, July 2). Search engine features for Webmasters [On-line] <http://www.searchenginewatch.com/webmasters/features.html>
  25. Vaughan, Liwen & Thelwall, Mike. (2004). Search Engines Bias : Evidence and Possible Causes, Information Processing and Management, Vol. 40, No. 4, pp. 693-707.
  26. Unerwood, Lee. A Brief History of Search Engines – Webreference.com. September, 2003. Retrieved from the WWW at May 20, 2005.
  - the light of 1998. Journal of Documentation, 54, (3), 1998, pp. 284-29.
  16. Madar Research Group (2004). NUA Internet Surveys : Arab World Set to Get Online. Retrieved from the Web at July, 25, 2004. <http://www.nua.ie/surveys/index.cgi>
  17. Moukdad, H. (2002). Language Based Retrieval of Web Documents : an Analysis of Arabic Recognition Capabilities of Two Major Search Engines. In proceedings of the 65<sup>th</sup> ASISIT Annual Meeting, Vol. 39 (ASISIT 2002), pp. 551-563.
  18. Mowshowitz, A., Kawaguchi, A. (2002). Assessing Bias in Search Engines. Information Processing and Management, 35 (4), pp. 443-462.
  19. Mohamed, Khaled (2004). Merging Multiple Search Results for Meta Search Engines. Ph.D. Dissertation, University of Pittsburgh, School of Information Science, 204 p.
  20. Notess, G.R. Measuring the Size of Internet Databases. Database (Weston, Conn.). V. 20 (October, November), pp. 69-70.
  21. Notess, G.R. (2004). Search Engines Statistics : Relative Size

- العربية على الإنترنت : دراسة تقييمية .  
 إشراف محمود عفيفي ، زين عبد الهادي ،  
 القاهرة . أطروحة ماجستير . جامعة حلوان ،  
 كلية الآداب ، قسم المكتبات والمعلومات .
- (٥) زين عبد الهادي ( أبريل ٢٠٠٢ ) .  
 محركات البحث على شبكة الإنترنت :  
 دراسة تجريبية مقارنة . مجلة المكتبات  
 والمعلومات العربية ، س٢٢ ، ع٢٤ ،  
 ص٥-٤٤ .
- (٦) سعد الهجرسي (٢٠٠٢) . مواقع الثقافة  
 العربية . دراسات عربية في المكتبات  
 والمعلومات . مج ٣ ، ع٢٤ ، ص٣٠-٤٥ .
- (٧) ضياء الدين عبد الواحد (٢٠٠٤) . واجهات  
 الاستخدام لنظم استرجاع المعلومات المتاحة  
 على شبكة الإنترنت : دراسة تقييمية .  
 إشراف سهير أحمد محفوظ ، زين عبد  
 الهادي . أطروحة ماجستير . جامعة حلوان ،  
 قسم المكتبات والمعلومات .
- (٨) محمد فتحى عبد الهادي (٢٠٠١) .  
 الإنترنت وخدمات المكتبات والمعلومات :  
 دراسة تحليلية فى ضوء الإنتاج الفكرى العربى .  
 المجلة العربية للمعلومات . ص٩٧-١٢٧ .

<http://www.webreference.com/authoring/search-history/>

27. UNICODE, INC, (2004). Standard Translation Arabic Unicode. <http://www.unicode.org>
28. Yahoo (2005). Yahoo Directory Tips. Retrieved from the Web at 2, May, 2005. [www.yahoo.com](http://www.yahoo.com)

### المراجع العربية :

- (١) أين (٢٠٠٥) . محرك البحث أين .  
 مسترجعة من الشبكة العنكبوتية فى نوفمبر  
 ٢٠٠٥ . <http://search.ayna.com>
- (٢) خالد رياض (١٩٩٩) . أدلة ومحركات  
 بحث شبكة الإنترنت : دراسة مقارنة فى  
 المؤتمر التاسع للاتحاد العربى للمكتبات  
 والمعلومات . تونس . المنظمة العربية للتربية  
 والثقافة والعلوم . ص١٣٧-١٤٥ .
- (٣) خالد عبد الفتاح محمد (١٩٩٩) . التعاون  
 والتنسيق بين مراكز المعلومات القطاعية  
 والشبكة القومية للمعلومات : دراسة ميدانية .  
 أطروحة ماجستير . جامعة القاهرة ، كلية  
 الآداب ، ص٢٤٥ .
- (٤) داليا نصار (٢٠٠٤) . محركات البحث

Internet Usage Statistics : Usage and Population Statistics (2005). World Internet Users and Population Statistics. Retrieved from the web at 20, November, 2005. Available online at : <http://www.internetworldstats.com/stats.htm>

December, 1995	16 millions	0.4 %	IDC
December, 1996	36 millions	0.9 %	IDC
December, 1997	70 millions	1.7 %	IDC
December, 1998	147 millions	3.6 %	C.I. Almanac
December, 1999	248 millions	4.1 %	Nua Ltd.
March, 2000	304 millions	5.0 %	Nua Ltd.
July, 2000	359 millions	5.9 %	Nua Ltd.
December, 2000	451 millions	7.4 %	Nua Ltd.
March, 2001	458 millions	7.6 %	Nua Ltd.
June, 2001	479 millions	7.9 %	Nua Ltd.
August, 2001	513 millions	8.6 %	Nua Ltd.
April, 2002	558 millions	8.6 %	Internet World Stats
July, 2002	569 millions	9.1 %	Internet World Stats
September, 2002	587 millions	9.4 %	Internet World Stats
March, 2003	608 millions	9.7 %	Internet World Stats
September, 2003	677 millions	10.6 %	Internet World Stats
October, 2003	682 millions	10.7 %	Internet World Stats
December, 2003	719 millions	11.1 %	Internet World Stats
February, 2004	745 millions	11.5 %	Internet World Stats
May, 2004	757 millions	11.7 %	Internet World Stats
October, 2004	812 millions	12.7 %	Internet World Stats
December, 2004	817 millions	12.7 %	Internet World Stats
March, 2005	888 millions	13.9 %	Internet World Stats
June, 2005	938 millions	14.6 %	Internet World Stats
September, 2005	957 millions	14.9 %	Internet World Stats

