



تعد شبكة الإنترنت من الناحية العملية مخزناً ضخماً للمعلومات. متعدد التخصصات ومختلف اللغات. وقد برزت شبكة الإنترنت في أوائل التسعينيات كوسيلة سهلة ورخيصة للنشر في شتى المجالات خاصة مع تطور شبكة النسيج العالمي (World Wide Web- www)، حيث أصبحت الوسيلة المفضلة للنشر من قبل الجهات التجارية والحكومية والتعليمية وغيرها. وتقدم شبكة الإنترنت العديد من الخدمات مثل خدمات التراسل الإلكتروني وخدمات البحث ونشر المعلومات. وتقدم الشبكة بعض الأدوات التي توفر إمكانيات استخدام الخدمات المتوفرة عليها، ومن أشهر تلك الأدوات مجموعة من محركات البحث التي تستخدم لتحديد أماكن وجود المعلومات على شبكة الإنترنت والوصول إليها.

موضوعاتها وأماكن نشرها وملخصات محتوياتها. ويكون الدليل في الغالب دقيقاً في تصنيف المعلومات ولكنه لا يعطي رصداً لكامل محتويات الإنترنت، لاعتماده على التحديث اليدوي الذي يكون في الغالب مكلفاً من حيث الجهد والوقت. ومن أشهر الأدلة على شبكة الإنترنت دليل ياهو (Yahoo) الذي أنشئ في أواخر عام ١٩٩٤م ويحوي رصداً لأكثر من مليون موقع، ويوجد منه نسخ متعددة بلغات مختلفة وفي أماكن متفرقة من العالم.

● فهرس الإنترنت

تعطي فهرس الإنترنت رصداً ألياً للتعابير والمفردات الواردة في المعلومات المنشورة على شبكة الإنترنت مع بيان بأماكن نشرها. وتتاح تلك الفهارس للمستخدم للبحث فيها عن مفردات أو تعابير باستخدام طرق بحث مختلفة. وبسبب طبيعتها الآلية تقوم الفهارس برصد معلومات أكثر من تلك المعلومات التي تحويها الأدلة، وتتميز بتغطية كبيرة تصل لمئات الملايين من صفحات النسيج.

وتعد محركات البحث المبنية على الفهارس من أشهر أنواع محركات البحث وأكثرها استخداماً، وتتميز بتغطية جغرافية واسعة، وسهولة في الاستخدام مع خصائص بحث متقدمة. ولكن يعاب عليها

الفترة عدد من محركات البحث. وتختلف تلك المحركات من حيث الخصائص والقدرات ولكنها تعتمد على تقنيات أساسية واحدة لتحديد أماكن نشر المعلومة وفهرستها وإتاحتها للمستخدمين عبر شبكة الإنترنت. يستعرض هذا المقال الأساسيات الفنية العامة المستخدمة في محركات البحث وطرق عملها وسبل استخدامها.

تصنيف محركات البحث

تتوفر على شبكة الإنترنت مجموعة كبيرة من محركات البحث يمكن تصنيفها حسب محتوياتها، والغرض منها، وطريقة ترتيب بياناتها و تصفحها والبحث فيها، ومن أشهر تلك الأنظمة ما يلي:

● أدلة الإنترنت

تقوم أدلة الإنترنت بتصنيف أغلب محتويات الإنترنت حسب موضوعاتها على شكل سرد هرمي مشابه للتصنيف المستخدم في المكتبات. ويكون التصنيف في الغالب يدوياً حيث يقوم العاملون على الأدلة بتتبع مواقع نشر المعلومات وتسجيل

مع توفر الكم الهائل والضخم من المعلومات على شبكة الإنترنت - بتخصصاتها المتعددة ولغاتها المختلفة وأماكن نشرها المتباعدة - أصبح من الصعوبة على مستخدم شبكة الإنترنت الإحاطة بمكان نشر المعلومات بغية الوصول إليها. وقد برز خلال السنوات الثمان الأخيرة عدداً من أدوات وخدمات البحث التي تساعد مستخدم الإنترنت على تحديد أماكن نشر المعلومات، وذلك باستخدام محركات البحث التي هي أدوات تقوم بجمع وبناء وفهرسة المعلومات المتوفرة على شبكة الإنترنت لمساعدة مستخدمي الشبكة على البحث عنها والوصول لها، فضلاً عن مكان نشرها، فهي في هذه الحالة تعمل كوسيط بين المستفيد وناسر المعلومات.

تعد محركات البحث الوسيلة الأجدى والأفضل لتحديد مكان نشر المعلومات في موضوع محدد على شبكة الإنترنت. فمع الزيادة المطردة لكمية المعلومات المنشورة على شبكة الإنترنت العالمية وتعدد لغات نشرها وأساليب وطرق بثها برز خلال تلك

المعلومات والباحث عنها بالمزايا والخصائص التي يوفرها محرك البحث، حيث يهتم ناشر المعلومات بشهرة وكثرة الاستخدام، بينما يهتم الباحث عن المعلومات بالسرعة والدقة وصيانة وسرعة تحديث الفهارس، وحجم التغطية، وقوة وسائل البحث وإتاحة موقع محرك البحث. ويمكن تقسيم خصائص محركات البحث إلى مايلي:

● خصائص وطرق البحث

تعد خصائص وطرق البحث من أهم المزايا التي تهم المستخدم لمحركات البحث التي تختلف باختلاف الطرق التي تتيحها للمستخدم للبحث في فهارسها حيث أنه ليس هناك طريقة مقننة لصياغة أمر البحث. ويبين شكل (١) واجهة بحث قياسية لأحد محركات البحث.

تستخدم محركات البحث الروابط المنطقية لتقييد نتيجة البحث أو توسيعها في عملية البحث المنطقي، وذلك بموجب ثلاثة روابط منطقية هي:

١- و (AND) المنطقية وتستخدم للبحث عن الصفحات التي تحوي كل المفردات الواردة في صيغة البحث، وتستثنى من ذلك الصفحات التي لا تحوي أياً من الكلمات أو تحوي إحداها.

٢- أو (OR) المنطقية وتستخدم لتوسيع عملية البحث وذلك بالبحث عن جميع صفحات النسيج التي تحوي إحدى أو كل الكلمات المعطاة في صيغة البحث.

العالمي ومجموعات النقاش ويحوي فهارس لأكثر من مائة مليون صفحة نسيج. أما محرك البحث قوقل (<http://www.google.com>) فيعد من أضخم محركات البحث على الإطلاق على شبكة

الإنترنت حيث تتعدى عدد صفحات النسيج المفهرسة فيه أكثر من بليون صفحة.

● فهرس الفهارس

يتوفر على الشبكة- بالإضافة إلى الأدلة والفهارس - أداة بحث تسمى فهرس الفهارس (Meta Search Engine) ، وهي لا تحوي فهارس بحد ذاتها ولكنها تستخدم الفهارس المتوفرة على الشبكة للبحث عن المعلومات المطلوبة من المستفيد. ويتميز فهرس الفهارس بواجهة تخاطبية سهلة، ونتيجة بحث شاملة، ولكن يعاب عليه بدائية البحث، ورداءة نتائجه على وجه العموم.

خصائص محركات البحث

تتوفر على شبكة الإنترنت المئات من محركات البحث التي تعتمد على تقنية المفهرسة الآلية، وتختلف تلك المحركات من حيث الطريقة التي تتعرف بها على أماكن نشر المعلومات، وطرق فهرستها، وطريقة عرضها لنتائج البحث. ويهتم ناشر



● دليل ياهو في شبكة الإنترنت.

رداءة نوعية المعلومات المتاحة في الفهارس مقارنة بالمعلومات الموجودة في الأدلة، وعدم وجود واجهه تخاطبية معيارية.

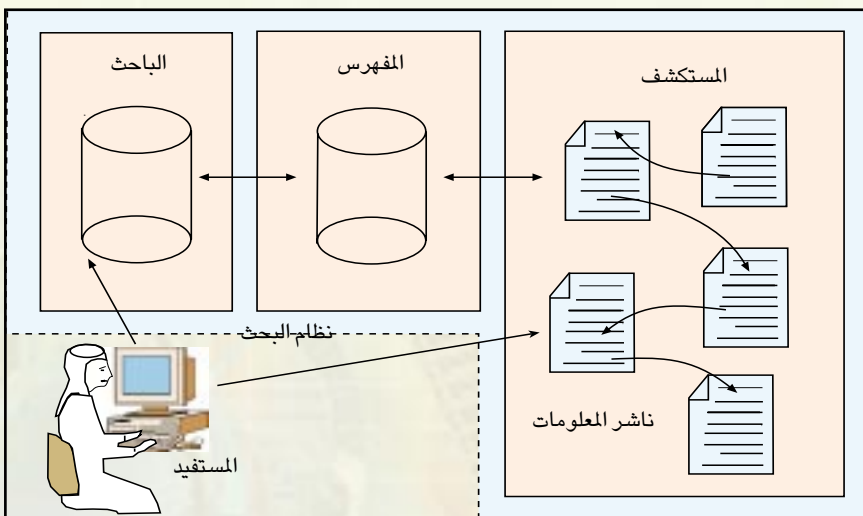
تتكون محركات البحث المعتمدة على المفهرسة الآلية من ثلاثة مكونات رئيسية، هي:-

● **المستكشف (Robot, Spider, Crawler)**، ويقوم بتتبع أماكن نشر المعلومات المسجلة في قاعدة بيانات المواقع أو عن طريق تتبع الروابط المدمجة في صفحات النسيج المترابطة، وجلب المعلومات المنشورة فيها للمفهرس.

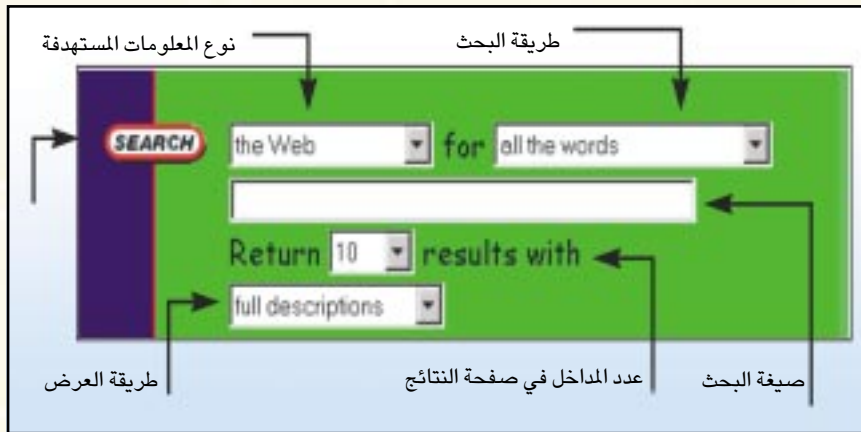
● **المفهرس**، ويقوم باستخدام كل أو معظم الكلمات الواردة في صفحة النسيج أو ملخصها وإدراجها في فهرس محرك البحث مع بعض الإشارات لموقعها النسبي في الصفحة وعنوان الصفحة نفسها، وتعتمد المفهرسة على تقنيات مقننة مبنية على نظام البحث في النصوص الكاملة.

● **الباحث**، ويشكل الواجهة التخاطبية بين المستخدم ومحرك البحث حيث يقوم باستقبال طلبات البحث عن المعلومات من المستفيد والبحث عنها في الفهرس وإشعار المستفيد بأماكن نشر تلك المعلومات. وتسمح معظم محركات البحث للمستخدم بالبحث بطرق مختلفة وعرض نتائج البحث بأشكال متعددة.

ويتوفر العديد من محركات البحث المعتمدة على الفهرس، من أشهرها قوقل (google) و **التافيسستا (altavista)** و **أل ذا وب (alltheweb)**. حيث يعد محرك **التافيسستا (http://www.altavista.com)** الذي أنشئ في نهاية عام ١٩٩٥ من قبل شركة **ديجيتال** من أشهر تلك الفهارس على الإطلاق، ويقوم بفهرسة صفحات شبكة النسيج



● مكونات محرك البحث.



● شكل (١) واجهة بحث قياسية.

وتعتمد حداثة الفهرس على طريقة المستكشف في جلب المعلومات. وتتراوح حداثة المعلومات في الفهرس بين دقائق وأشهر. ويعد إظهار تاريخ جلب الوثيقة من الموقع عند عرض نتائج البحث ميزة جيدة في بعض محركات البحث، لما تعطيه للمستخدم من معرفة بمدى حداثة المعلومات في فهرس محرك البحث.

تتعرف محركات البحث على مواقع وجود المعلومات على شبكة الإنترنت وفهرستها بطريقتين. الأولى عن طريق إشعار ناشر الموقع عنوان موقعه في مكان مخصص لدى محرك البحث، والأخرى عن طريق تتبع محرك البحث للوصلات المؤدية للموقع.

تحديد خصائص الاستكشاف طرق
استكشاف المعلومات على شبكة النسيج العالمي وحجمها ومحتواها. ومن هذه الخصائص ما يلي:

● **خاصية التغطية**، وتحدد نوع ومصدر المعلومات الموجودة في فهرس محرك البحث وتشمل - في الغالب - صفحات النسيج، وحلقات النقاش، والأدلة البيضاء والصفراء.

● **خاصية المحتوى**، وتحدد حجم النص الفهرس من قبل محرك البحث سواء كان بفهرسة النص الكامل للصفحة المنشورة أو بفهرسة ملخص الصفحة فقط.

● **خاصية الحجم**، وتتحكم في عدد الصفحات الفهرسة من قبل محرك البحث، حيث يقاس الحجم بملايين الصفحات.

● **خاصية العمق**، وتعمل على قياس عدد الصفحات وعمقها في الموقع بداية من

● **خاصية البحث النطاقي**، وتسمح بحصر البحث في نطاق شبكي محدد مثل النطاق التعليمي أو التجاري أو منطقة جغرافية مثل السعودية.

وعند البحث باللغة الإنجليزية تقوم بعض محركات البحث بالتفريق بين الأحرف الكبيرة والصغيرة بينما لا تعطي معظم محركات البحث أهمية لاختلاف الأحرف وتقوم بعرض نتائج شاملة. تمكن معظم محركات البحث من دمج نوع أو أكثر من أنواع البحث في صيغة بحث واحدة وتستخدم الأقواس عادة لتحديد أولويات التنفيذ.

● خصائص الاستكشاف

يمكن للمستكشف التعرف على مدى أهمية وشعبية صفحة النسيج المنشورة بالتعرف على عدد المواقع المرتبطة بها، ويستخدم تلك المعلومات لتحديد جدوى فهرسة صفحة النسيج ومكان عرضها في صفحة نتائج البحث. وتتمكن بعض محركات البحث التعرف على دورة تحديث الموقع، ومنها يمكنها جدولة زيارات المستكشف لتحديث بيانات الموقع في الفهارس الخاصة بمحرك البحث. يمكن للموقع المراد فهرسته التحكم في المستكشفات "الصدقية" ومنعها من فهرسة الموقع أو السماح لها بفهرسة بعض أو كل الموقع بإدراج التعليمات في ملف (robots.txt) أو من خلال عنصر الترميز (meta tag) في صفحة النسيج.

تتمايز محركات البحث بمدى حداثة المعلومات الموجودة في فهرسها وكمية المعلومات التي تقوم بفهرستها يومياً.

٢- غير (NOT) المنطقية، وتستخدم لتقييد عملية البحث بالبحث عن جميع الصفحات التي تحوي الكلمة الأولى، ويستثنى من ذلك الصفحات التي تحوي الكلمة الثانية.

تسمح بعض محركات البحث باستخدام اللغة الطبيعية في صيغة البحث وذلك بإزالة الكلمات غير الدالة واستخدام إحدى وسائل البحث في البحث عن باقي الكلمات الواردة في صيغة البحث. كما تقدم بعض المحركات إمكانية البحث عن المفاهيم حيث يقوم محرك البحث بتحليل محتويات الصفحات أثناء عملية الفهرسة والربط بين الكلمات والتعابير فيها. وأثناء عملية البحث يقوم المحرك بالبحث عن الكلمات المطلوبة والكلمات المرتبطة بها، وإبراز النتائج في مجموعة واحدة. وذلك حسب الخصائص التالية:

● **خاصية البحث عن التعابير**، وتسمح بالبحث عن صفحات تحتوي على تعابير محددة، وتستخدم عادة في ذلك علامة التنصيص " " لإدراج التعابير المطلوب البحث عنها.

● **خاصية البحث المكاني**، وتستخدم للبحث عن صفحات تحتوي على كلمات محددة يفصلها عدد محدد من الكلمات.

● **خاصية تحديد لغة النشر**، وتسمح في بعض محركات البحث إمكانية حصر البحث في صفحات منشورة بلغة محددة.

● **خاصية البحث بالتر**، وتمكن من البحث عن صفحات تحوي مفردات بغض النظر عن طريقة تهجئة المفردات أو اللواحق المرتبطة بها. وتزيد هذه الطريقة من كمية المعلومات المسترجعة وكذلك من المعلومات التي لا تتعلق بطلب المستفيد.

● **خاصية البحث عن وسائط متعددة**، ويستفاد منها في البحث عن معلومات غير نصية في شبكة النسيج كالصور والتسجيلات الصوتية والمقاطع الفلمية.

● **خاصية البحث الحقلي**، وتستخدم لحصر نتائج البحث في أحد حقول صفحة النسيج أو المعلومات عنها مثل العنوان أو المتن أو المصدر.

● **خاصية البحث الزمني**، وتسمح بحصر البحث في فترة زمنية محددة.

محركات البحث

هذه الخصائص كثرة الاشتقاق والمترادفات واستعمال المفاهيم المركبة في النصوص العربية إضافة للاختلاف في طرق كتابة المصطلحات الأجنبية.

يتسبب كثرة الاشتقاق في تشتت المداخل في فهارس محركات البحث وذلك لكثرة السوابق التي ترتبط بالمفردة الواحدة مما يعقد عملية البحث ويسبب ضياع بعض المفردات في الفهارس ويجعلها - عملياً - غير قابلة للبحث. وتسبب كثرة الاشتقاقات - كذلك - زيادة في عدد المداخل في الفهارس بسبب السوابق واللواحق التي ترتبط بالمفردة الواحدة، مما يجعل عملية البحث أبطأ وأكثر تعقيداً، أما كثرة المترادفات في اللغة العربية وتؤدي إلى تشتت المداخل والتعابير ذات المفهوم الواحد في الفهارس ومن ثم إلى صعوبة عملية البحث.

يغلب في اللغات الطبيعية استخدام المفاهيم المركبة التي تتكون من مفردتين أو أكثر، ونظراً لطرق التعبير المختلفة والمتعددة في اللغة العربية فإن البحث الحر عن تلك التعابير يؤدي لاسترجاع القليل من المعلومات ذات العلاقة، ولاسترجاع المفاهيم المركبة تستخدم تقنيات بحث مختلفة مثل البحث باستخدام الروابط المنطقية أو البحث المكاني. كذلك يكثر في الكتابة باللغة العربية استخدام المصطلحات الأجنبية المكتوبة بأحرف عربية والتي وتختلف طرق كتابة من مؤلف إلى آخر ومن بلد إلى آخر. عليه فمن الصعب البحث عن المفردات ذات المفهوم الواحد.

بالإضافة إلى ذلك تواجه محركات البحث العربية على شبكة الإنترنت بعض المشاكل التي تؤثر على كمية المعلومات المفهرسة وطريقة البحث عنها، فمنها على سبيل المثال اختلاف شفرة المحارف، فمع وجود أكثر من شفرة محارف عربية وعدم وجود مواصفة قياسية إلزامية، درج ناشرو المعلومات العربية على شبكة الإنترنت على نشر المعلومات بشفرات محارف مختلفة، مما جعل التعرف على محتويات صفحة النسيج وفهرستها ليس أمراً سهلاً. وقد تم تطوير أدوات لمساعدة محركات تنظم البحث على التعرف على لغة نشر المعلومات والمحارف المستخدمة لذلك تمهيداً لفهرستها.

عنها عند القيام بعملية البحث، حيث أن حذفها يقلل حجم الفهرس ويسرع عملية البحث.

● خصائص عرض نتائج البحث

تختلف محركات البحث في طرق عرض نتائج البحث، فليس هناك طريقة ثابتة وقياسية لعرض المعلومات وتتكون عناصر العرض من: العنوان والملخص وعدد السجلات في الصفحة الواحدة، ويبين شكل (٢) نموذجاً لنتيجة بحث لأحد محركات البحث.

تعطي بعض محركات البحث أهمية كبيرة لمكان وجود المفردات وتكرارها في صفحة النسيج إمكانية عرض الصفحة ضمن نتائج البحث وترتيبها في صفحة نتائج البحث. كما تعطي بعض النظم أهمية لشعبية الصفحة ولوجودها ضمن الدليل المرتبط بمحرك البحث. وتتمايز محركات البحث في طريقة **عنوان السجل** ووصفه عند عرضه ضمن نتائج عملية البحث، وكذلك في عدد السجلات المدرجة في صفحة نتائج البحث الواحدة.

شبكة الإنترنت واللغة العربية

تختص نصوص اللغة العربية بخصائص تجعل عملية التعرف عليها وعملية الفهرسة الآلية لها والبحث فيها أمراً بالغ التعقيد مقارنة باللغات اللاتينية، ومن

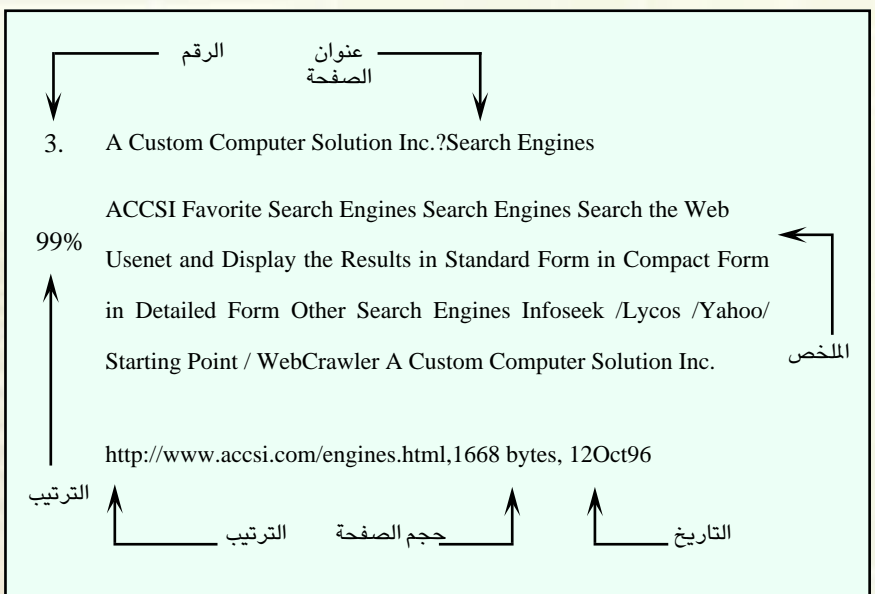
الصفحة المعلنة أو الصفحة التي يقوم المستكشف بزيارتها وفهرستها. وقد يحد من عدد الصفحات المفهرسة استخدام الأطر أو الخرائط الصورية. ويكمن لصاحب الموقع - في الغالب - التحكم في تعاطي المستكشف لبيانات الموقع ولذلك يعطي محرك البحث للمستكشف إسماً يعرف به يستخدم من قبل صاحب الموقع لإعطاء التعليمات لمستكشفات "الصديقة".

* **خاصية الإدراج المبرمج**، وتبين متوسط الوقت اللازم لإضافة الموقع أو الصفحة للفهرس بعد تسجيلها لدى المستكشف من قبل الناشر أو المستفيد.

* **خاصية الإدراج غير المبرمج**، وتحدد متوسط الوقت اللازم لإضافة الموقع أو الصفحة للفهرس من قبل المستكشف بعد التعرف على مكان الموقع عن طريق تتبع الترابط بين الصفحات الذي يتطلب - في الغالب - وقتاً أطول.

● خصائص الفهرسة

تحدد خصائص الفهرسة الطرق المتبعة من محركات البحث لإنشاء الفهارس، ومن أهم تلك الخصائص قدرة محرك البحث على التعرف على عناصر التوصيف المدرجة عن طريق عنصر الترميز " ميتا " (Meta tag) - إعطاءه أهمية ووزن أثناء عملية الفهرسة والبحث - القدرة على تمييز **كلمات الوقف** وهي الكلمات غير الدالة من الفهرس وحذفها وعدم دعم عملية البحث



● شكل (٢) نتيجة بحث قياسية.