

محركات البحث

Search engines

المحاضر: حسام يونسو

بكالوريوس تقنية معلومات – تخصص تقنية

- إدراك التحديات الحديثة في البحث
- الخاتمة والأسئلة
- مراجع علمية للمادة
- المخرجات المتوقعة من الدرس
- المقدمة
- فهم أهمية محركات البحث
- التعرف على تاريخ محركات البحث وتصفح الويب
- معرفة مكونات الويب
- استيعاب كيفية عمل محركات البحث
- التمييز بين البحث والتصفح

المخرجات المتوقعة من الدرس

بنهاية هذا الدرس، يتوقع من الطلاب فهم أهمية محركات البحث في حياتهم اليومية، وكيف تسهم في تسهيل الوصول إلى المعلومات في مجالات متعددة مثل التعليم والصحة والتكنولوجيا. سيتعرف الطالب على تاريخ محركات البحث وتصفح الويب منذ البداية وصولاً إلى تطوير الويب وأثر ذلك على تجربة المستخدم. كما سيتعلمون الفرق بين البيانات المنظمة وغير المنظمة، بالإضافة إلى فهم كيفية عمل محركات البحث من خلال الزحف والفهرسة والاستعلام، ومعرفة مفهوم PageRank. سيتمكن الطالب من التمييز بين البحث باستخدام الكلمات المفتاحية والتصفح عبر الروابط، مع فهم التحديات التي تواجه محركات البحث مثل الكم الهائل من البيانات والمحفوظ غير المرغوب. علاوة على ذلك، سيتعرفون على مستقبل البحث من خلال التعلم الآلي والبحث المخصص والبحث الاجتماعي وكيفية تحسين نتائج البحث. وأخيراً، سيتمكنون من استخدام محركات البحث بفعالية أكبر وتحقيق دقة أفضل في النتائج، والاستعداد لمواكبة التطورات المستقبلية في تكنيات البحث.

محركات البحث أصبحت جزءاً لا يتجزأ من حياتنا اليومية، حيث نستخدمها بشكل مستمر للوصول إلى المعلومات في مختلف المجالات مثل التعليم، الصحة، والتكنولوجيا. مع تطور الإنترنت، أصبحت هذه المحركات أداة أساسية تسهل علينا العثور على ما نحتاجه في ثوانٍ معدودة. في البداية، كان البحث عن المعلومات يتطلب التوجه إلى المكتبات أو الاستعانة بالخبراء، لكن الآن أصبح بإمكان أي شخص الوصول إلى كم هائل من المعلومات عبر محركات البحث مثل Google وBing. في هذا الدرس، سنتعرف على كيفية عمل محركات البحث، تاريخ تطورها، وأساسيات البحث والتصفح، بالإضافة إلى التحديات التي تواجهها هذه المحركات في العصر الحالي، مع تسلية الضوء على تقنيات المستقبل التي ستساهم في تحسين نتائج البحث.

أهمية محركات البحث في حياتنا اليومية

حركات البحث أصبحت جزءاً لا يتجزأ من حياتنا اليومية لأنها تسهل الوصول إلى المعلومات بسرعة وفعالية، بغض النظر عن المجال الذي نبحث فيه.

مثال في التعليم: إذا كنت تدرس موضوعاً في الرياضيات، يمكنك البحث عن "كيفية حل معادلة الدرجة الثانية" وستحصل على شروحات ونماذج تمارين.

أهميتها العامة:

- توفير الوقت والجهد.
- تحسين الوصول إلى المعرفة.
- مساعدة الأفراد في اتخاذ قرارات مستنيرة.
- تعزيز الإنتاجية في مختلف جوانب الحياة.



مساهمة التقنية في تغيير طريقة وصولنا إلى المعلومات

• قديماً:

الاعتماد الأساسي كان على المكتبات الورقية للحصول على المعرفة.

على سبيل المثال: البحث عن كتاب محدد كان يتطلب زيارة مكتبة، العثور عليه بين الرفوف، وربما الانتظار إذا كان مجازاً.

• الآن:

بسبب محركات البحث، أصبحت المعلومات في متناول اليد لأي شخص لديه اتصال بالإنترنت.

يمكن إجراء بحث بسيط عن موضوع ما والحصول على آلاف النتائج خلال ثوانٍ.

مثال: كتابة سؤال مثل "ما هي أعراض نزلات البرد؟" يعرض مقالات طبية، فيديوهات، وتجارب شخصية.

تاريخ محركات البحث وتصفح الويب



البداية:

كانت البداية مع Vannevar Bush ومفهوم الـ Memex عام

1945

- **Memex** كان جهازاً تخلياً يشبه مكتبة شخصية تحتوي على معلومات يمكن ربطها بعضها البعض.
- الفكرة كانت أساساً لمفهوم الروابط التشعبية (Hyperlinks) التي نستخدمها الآن.

تاريخ محركات البحث وتصفح الويب

تطور الويب:



- تطوير "الويب" على يد **Tim Berners-Lee** عام **1990**
- اخترع الويب لربط المستندات عبر الإنترنت باستخدام:
 - عناوين المواقع (URL).
 - بروتوكول نقل النص التشعبي (HTTP).
 - لغة HTML لعرض الصفحات.

إطلاق أول متصفح ويب مثل Mosaic عام 1993

• Mosaic كان أول متصفح يوفر واجهة مستخدم رسومية.

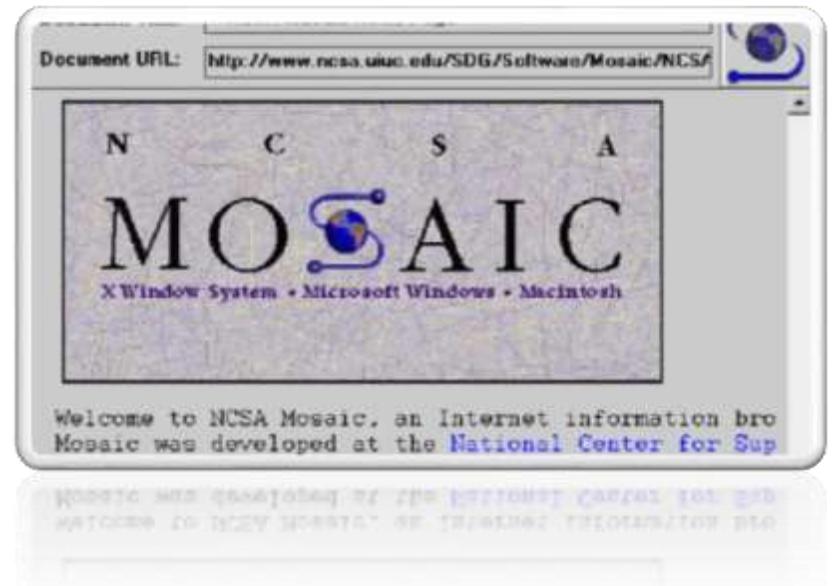
• جعله أكثر سهولة للناس لتصفح الويب.

تطور محركات البحث

• (1994) Yahoo: بدأت كدليل يدوياً يصنف المواقع.

• (1998) Google: قدمت طريقة مبتكرة لترتيب النتائج بناءً

على شعبيتها وروابطها.



- 1- كيف كانت الطريقة التقليدية للحصول على المعلومات قبل ظهور محركات البحث؟
- 2- ما الفرق الذي أحدثته التقنيات الحديثة في طريقة الوصول إلى المعلومات؟
- 3- ما هو مفهوم الـ **Memex** الذي اقترحه **Vannevar Bush** الذي ابتكره؟
- 4- ما هي التقنيات الثلاث التي ابتكرها **Tim Berners-Lee** لإنشاء الويب؟



تعريف الويب

الويب هو شبكة افتراضية ضخمة تربط بين مiliarات الصفحات والواقع عبر الإنترنت. الهدف الأساسي للويب هو توفير وسيلة للوصول إلى المعلومات بطريقة سريعة وسهلة، حيث يتم توصيل المحتوى عبر بروتوكولات معينة (مثل HTTP) ولغات تصميم (مثل HTML). هذه الشبكة ليست محدودة بموقع جغرافي، بل تربط العالم بأسره.

البيانات المنظمة (Structured Data)

Structured data

ID	Name	Age	Degree
1	John	18	B.Sc.
2	David	31	Ph.D.
3	Robert	51	Ph.D.
4	Rick	26	M.Sc.
5	Michael	19	B.Sc.

- هي بيانات مرتبة ومخزنة بطريقة يمكن فهمها بسهولة بواسطة الأنظمة مثل قواعد البيانات.
- يتم تقديم البيانات بشكل جدولي أو منظم (مثل ملفات Excel أو قواعد بيانات SQL).
- مثال: قائمة طلاب تحتوي على الأسماء وأرقام الهواتف وعناوين البريد الإلكتروني.

البيانات غير المنظمة (Unstructured Data)

- هي بيانات متنوعة وغير مرتبة، ولا يمكن تخزينها مباشرة في جداول منتظمة.
- تشمل نصوصاً، صوراً، مقاطع فيديو، أو صفحات ويب مليئة بمحفوظات متنوعة.
- مثال: صفحة إخبارية تحتوي على مقالات، روابط، وصور مدمجة مع النصوص.

Unstructured data

The university has 5600 students.

John's ID is number 1, he is 18 years old and already holds a B.Sc. degree.

David's ID is number 2, he is 31 years old and holds a Ph.D. degree. Robert's ID is number 3, he is 51 years old and also holds the same degree as David, a Ph.D. degree.

الويب العميق (Deep Web)

هو جزء من الويب غير مرئي لمحركات البحث التقليدية مثل Google أو Bing. السبب في عدم ظهوره هو أن هذه المحتويات مخزنة في قواعد بيانات محمية، أو تحتاج إلى تسجيل دخول للوصول إليها.

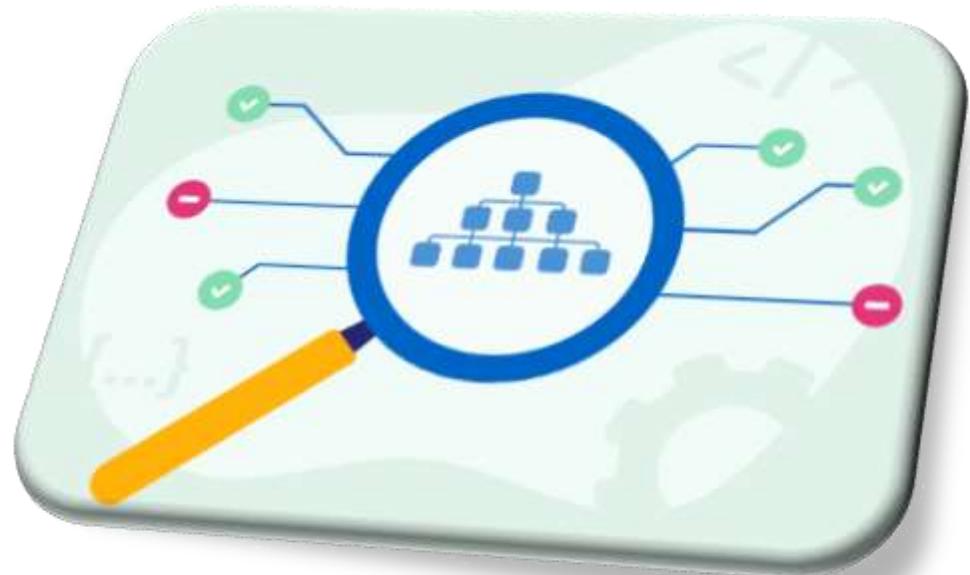
أمثلة على المحتوى:

- السجلات الطبية الخاصة بالمرضى المخزنة في قواعد بيانات المستشفيات.
- الأبحاث الأكاديمية أو المقالات العلمية التي تكون محمية خلف جدران دفع أو تسجيل دخول.
- بينما يمكن لمحركات البحث الوصول إلى المحتوى العادي على الويب (مثل الموقع الإخبارية أو المدونات)، فإن الويب العميق يحتوي على كمية هائلة من البيانات المخفية، والتي قد تكون ذات أهمية كبيرة ولكنها غير متاحة بسهولة.

كيفية عمل محركات البحث

محركات البحث تعتمد على عملية من ثلاثة خطوات رئيسية لجمع المعلومات وتنظيمها وعرضها للمستخدمين:

- الزحف (Crawling)
- الفهرسة (Indexing)
- الاستعلام (Querying)



الزحف (Crawling)



- يتم استخدام برمجيات تسمى "العناكب" (Spiders) أو "الروبوتات" (Bots) لزيارة المواقع الإلكترونية بشكل آلي.
- الهدف هو جمع البيانات من صفحات الويب، بما في ذلك النصوص، الصور، والروابط.
- مثال عملي: إذا أنشئت مدونة جديدة عن "التكنولوجيا"، ستزور العنكبوت المدونة وتفهرسه ليصبح متاحاً للبحث.

- بعد جمع البيانات، يتم تخزينها في قاعدة بيانات كبيرة تسمى "الفهرس".
- الفهرس ينظم المحتوى بطريقة تجعل محركات البحث تسترجع المعلومات بسرعة عند الحاجة.
- مثال: إذا كانت الصفحة تحتوي على كلمة "رياضية"، يتم إدخالها في قسم الفهرس الخاص بالرياضة ليسهل العثور عليها.



الاستعلام (Querying)

- عند كتابة استعلام (مثل "أفضل مطاعم قرية مني")، تستخدم محركات البحث الفهرس للعثور على الصفحات التي تحتوي على معلومات ذات صلة.
- تظهر النتائج بناءً على عوامل مثل الموقع الجغرافي وتقييمات المستخدمين.
- مثال: عند البحث عن "أفضل مطعم"، ستظهر النتائج الأقرب جغرافياً مع ترتيب يعتمد على التقييمات.



PageRank هو نظام ابتكرته Google لتقدير الصفحات على الإنترنط. يعتمد النظام على فكرة أن الصفحة التي تحصل على العديد من الروابط من صفحات موثوقة هي أكثر أهمية. كيف يعمل؟

- كل رابط يُعتبر بمثابة "تصويت" لصالح الصفحة المرتبطة.
- الجودة أيضاً مهمة: الروابط من المواقع ذات المصداقية العالية تعطي وزناً أكبر.

مثال: إذا كانت صفحة تتحدث عن "التصميم الجرافيكي" وتشير إليها موقع موثوقة مثل جامعات أو شركات عالمية، فإن ترتيبها سيكون أعلى في نتائج البحث مقارنة بصفحة مشابهة ولكن لا تشير إليها سوى موقع قليلة وغير معروفة.

محركات البحث تُسطّح الوصول إلى المعلومات عبر تنظيم مiliarات الصفحات بذكاء، مع إعطاء الأولوية للمحتوى الأكثر صلة وجودة.

- 1- ما هي الخطوات الثلاث الرئيسية التي تعتمد عليها محركات البحث في عملها؟
- 2- ما دور العناكب (Spiders) في عملية الزحف؟
- 3- كيف يساعد نظام PageRank في ترتيب الصفحات في نتائج البحث؟
- 4- لماذا تعتبر جودة الروابط أكثر أهمية في PageRank من مجرد عددها؟

- **المفهوم:** عملية تعتمد على إدخال كلمات مفاتيحية في محرك البحث للوصول إلى صفحات معينة تتعلق بموضوع البحث.
- **كيف يعمل:** عندما يكتب المستخدم استعلاماً، يقوم محرك البحث بمطابقة الكلمات المفاتيحية مع المحتوى في قاعدة بياناته (الفهرس) لتحديد الصفحات ذات الصلة.
- **مثال:** إذا كنت تبحث عن "أفضل الكتب التقنية"، سيُظهر محرك البحث قائمة من النتائج التي تحتوي على توصيات وراجعات لكتب تقنية.



التصفح (Browsing)



- **المفهوم:** عملية تعتمد على التنقل بين الروابط الموجودة في صفحات الويب لاكتشاف محتوى جديد.
- **كيف ي عمل:** المستخدم يبدأ من صفحة معينة، ويستخدم الروابط (Hyperlinks) للانتقال إلى صفحات أخرى.
- **مثال:** الدخول إلى صفحة عن "البرمجة" ثم التنقل إلى صفحة تتحدث عن "تصميم ويب" عبر رابط في الصفحة الأولى.



الدقة (Precision):

- **المفهوم:** نسبة النتائج التي تتعلق بموضوع البحث مقارنة بالنتائج الإجمالية.
- **المشكلة:** قد تظهر نتائج غير دقيقة أو غير مفيدة ضمن القائمة.

مثال على المشكلة: عند البحث عن "لغة البرمجة بايثون"، قد تظهر نتائج عن الثعبان "Python".



التغطية (Recall):

- **المفهوم:** نسبة الصفحات ذات الصلة التي تظهر في نتائج البحث مقارنة بجميع الصفحات ذات الصلة الموجودة على الويب.
- **المشكلة:** بعض الصفحات المهمة قد لا تظهر بسبب محدودية فهرسة محرك البحث أو ترتيب النتائج.
- **مثال على المشكلة:** عند البحث عن "تصميم تطبيقات"، قد لا تظهر صفحات جديدة أو أقل شهرة على الرغم من أنها مفيدة.

التحديات الحديثة في البحث

الكم الهائل من البيانات

التحدي:

- محركات البحث تواجه صعوبة كبيرة في التعامل مع الكم الضخم من المحتوى على الإنترنط.
- يجب على المحركات تنظيم وفهرسة ملايين الصفحات باستمرار وتحديث قواعد البيانات الخاصة بها.

سبب المشكلة:

- الإنترنط ينمو بشكل يومي، مع إضافة صفحات جديدة وتحديث أخرى.
- الأخبار والمقالات، على سبيل المثال، يتم تحديثها أو تغييرها باستمرار.

مثال عملي: عند البحث عن "آخر أخبار الطقس"، تحتاج محركات البحث إلى التأكد من أن النتائج محدثة بالبيانات الأحدث.

التحديات الحديثة في البحث

إدارة البيانات الديناميكية

التحدي:

- بعض البيانات تحتاج إلى تحديث لحظي، مثل أسعار العملات، الأسهم، أو الأحداث الجارية.
- يجب أن تكون محركات البحث قادرة على عرض المعلومات المحدثة فور تغييرها.

سبب المشكلة:

- هذه البيانات تعتمد على مصادر ديناميكية وتتغير بشكل مستمر.

مثال عملي:

- البحث عن "سعر الدولار مقابل اليورو" يتطلب إظهار أحدث القيم وليس البيانات القديمة.

التحديات الحديثة في البحث

المحتوى غير المرغوب (Spam):

التحدي:

- هناك صفحات مضللة تهدف إلى جذب الزوار بطرق خادعة، دون تقديم قيمة حقيقية.
- هذه الصفحات قد تحتوي على عناوين مثيرة لكنها تقدم محتوى ضعيف أو غير مفيد.

سبب المشكلة:

- بعض المواقع تُستخدم لتحقيق أرباح غير مشروعة عن طريق الإعلانات أو بيع معلومات مغلوطة.

مثال عملي:

- موقع تعد بتقديم "وصفات تخسيس سحرية" لكنها تحتوي على معلومات غير علمية بهدف جذب الزوار.

كويزات (3)

3- ما هي المشكلة التي قد تواجه محركات البحث عندما تظهر نتائج غير دقيقة؟

- (A) التغطية.
- (B) الدقة.
- (C) التحديث المستمر للبيانات.
- (D) الخوارزميات المتغيرة.

4- ما هو المثال الذي يوضح مشكلة التغطية في البحث؟

- (A) نتائج بحث تحتوي على معلومات قديمة.
- (B) ظهور نتائج غير دقيقة عن "لغة البرمجة بايثون" تتعلق بالثعبان.
- (C) عدم ظهور صفحات مهمة في نتائج البحث رغم ارتباطها بالموضوع.
- (D) تنقل المستخدم بين الصفحات بشكل غير منظم.

1- ما هي عملية البحث في محركات البحث؟

- (A) إدخال استعلامات للحصول على نتائج غير دقيقة.
- (B) التنقل بين الروابط لاكتشاف محتوى جديد.
- (C) إدخال كلمات مفتاحية للوصول إلى صفحات معينة.
- (D) استخدام أدوات للتحقق من المحتوى المتعلق.

2- ما هو التحدي الرئيسي في إدارة البيانات الديناميكية في محركات البحث؟

- (A) عدم قدرة محركات البحث على التعامل مع النصوص العميقية.
- (B) التحدي في تحدث البيانات بشكل لحظي، مثل أسعار العملات.
- (C) التنقل بين الروابط لاكتشاف محتوى جديد.
- (D) ظهور نتائج غير دقيقة في البحث.

الأجوبة (كويز 1)

- 1- كان الاعتماد على المكتبات الورقية والخبراء للحصول على المعلومات.
- 2- وفرت وصولاً سريعاً وواسع النطاق إلى المعلومات عبر الإنترنـت.
- 3- جهاز تخيلي يشبه مكتبة شخصية تحتوي على معلومات يمكن ربطها ببعضها البعض.
- 4- عناوين المواقع (URL)، بروتوكول HTTP، ولغة HTML.

الأجوبة (كويز 2)

- 1- الزحف (Crawling)، الفهرسة (Indexing)، الاستعلام (Querying).
- 2- زيارة المواقع الإلكترونية وجمع البيانات مثل النصوص، الصور، والروابط لتكون جاهزة للفهرسة.
- 3- يقوم بتقييم الصفحات بناءً على عدد وجودة الروابط التي تشير إليها.
- 4- لأن الروابط من المواقع ذات المصداقية العالية تحمل وزناً أكبر عند تقييم الصفحة.

الأجوبة (كويز 3)

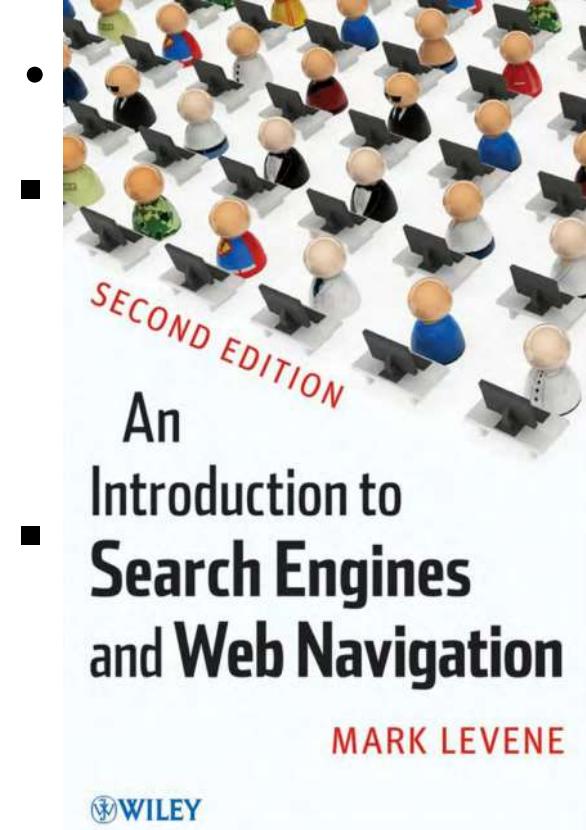
- 1- الإجابة: إدخال كلمات مفاتيحية للوصول إلى صفحات معينة.
- 2- الإجابة: التحدي في تحديث البيانات بشكل لحظي، مثل أسعار العملات.
- 3- الإجابة: الدقة.
- 4- الإجابة: عدم ظهور صفحات مهمة في نتائج البحث رغم ارتباطها بالموضوع.

اسم الكتاب: AN INTRODUCTION TO SEARCH ENGINES AND WEB NAVIGATION

رابط الكتاب الإلكتروني: <https://srikarthiks.wordpress.com/wp-senigne-hcraes-ot-noitcudortni-na-4t/2016/07/content/uploads/>

nd.pdf2-noitagivan-bew-dna

مؤلف الكتاب: Mark Levene



شكرا لكم