محركات البحث الدلالي في ظل تطبيقات الويب الدلالي

أ.د. فاتن سعيد بامفلح

قسم علم المعلومات

جامعة الملك عبدالعزيز

مرت الشبكة العنكبوتية بأجيال متعددة منذ أن تم تطويرها عام ١٩٩٠م على يد تيم بيرنـــرز لي Tim Berners- Lee ، فقد دعمت النشر الإلكتروني من خلال صفحات تســتخدم لغـــة ترميز النصوص الفائقة HTML ، ثم تطورت فيما بعد لتعتمد على النشر باستخدام نظم إدارة المحتوى. وقد كانت الشبكة العنكبوتية تحقق استرجاع المعلومات وتبادل المعرفة والمشـــاركة فيها بين المستفيدين من أي مكان دون اعتبار للحواجز الجغرافية.

ومع ظهور الجيل الثاني من الويب تم استخدام تقنيات وبروتوكولات حديثة للنشر الإلكتروني، وبدأت عملية استخدام الشبكة العنكبوتية تتجه نحو التفاعل بين المستفيد والنظام مـــن خـــلال التقنيات المختلفة مثل الويكي wiki ، و خدمة مـــوجز الويـــب (RSS) Really Syndication Simple ، والشبكات الاجتماعية، وغيرها... ومع الاتجاه نحو الجيل الثالث من الويب ظهر ما يعرف بتقنيات الويب الدلالي Semantic web ، الذي يعد ثورة في عالم المعلومات، حيـــث أحدث نقلة في معالجة المعلومات عن طريق تحقيق فهم الويب لمعاني ودلالات الكلمات التـــي يعالجها ويسترجعها؛ فلم يقتصر الأمر على ما كان عليه في السابق من التعامل مع الآلة مـــن خلال التعرف على اللغة الثنائية (صفر وواحد)، دون فهم لدلالات المكتوب عبر الويب.

وقد تبع هذا التطور، تطوراً في نظم استرجاع المعلومات المستخدمة مع هذا الجيل الجديد من الويب؛ حيث ظهر ما يعرف بالبحث الدلالي Semantic search ، ومحركات البحث التـــي تقدم هذا النوع من البحث.

وفي هذه الدراسة تعرف الباحثة بالبحث الدلالي وخصائصه، وما يميزه عن البحث التقليـــدي، مع دراسة تطبيقية تحليلية لاستخدام أحد محركات البحث الدلالي مقارنة بأحد محركات البحث التقليدي، حيث تغطي الدراسة كل من محرك بحث حكيا، ومحرك بحث جوجل، للتعرف على بعض خصائصهما البحثية والاسترجاعية خلال فترة إجراء الدراسة التطبيقية في شهر نوفمبر من عام ٢٠١٠م.

وتكمن أهمية هذه الدراسة في أنها تعد من الدراسات العربية القليلة التي تتجه نحو موضوع البحث الدلالي الذي يمثل ثورة حديثة في نظم استرجاع المعلومات. ويمكن أن تفيد هذه الدراسة في توجيه الاهتمام العربي بشكل أكبر نحو تطبيقات الجيل الثالث من الويب، واستثمار هذا التطور التقني في مجال البحث واسترجاع المعلومات في محركات البحث العربية عبر الشبكة العنكبوتية.

وتهدف هذه الدراسة إلى متابعة أبرز التطورات الحديثة في مجال استرجاع المعلومات، والتي نتجت عن ظهور الجيل الثالث من الويب . وينبثق من هذا الهدف الرئيس بعض الأهداف الفرعية المتمثلة في الآتي:

- إلقاء الضوء على مفهوم وخصائص البحث الدلالي الذي يمثل أحد الاتجاهات الحديثة في استرجاع المعلومات.
- فهم أبرز متطلبات عمل الويب الدلالي وتقنياته.
- التعرف على أبرز خصائص محركات البحث الدلالي مقارنة بمحركات البحث التقليدية.

وتجيب الدراسة على التساؤلات الآتية:

١. ما المقصود بالبحث الدلالي؟ وما متطلبات تطبيقه؟
٢. ما الفرق بين محركات البحث الدلالي ومحركات البحث التقليدية.

وتتبع الدراسة المنهج الوصفي عن طريق إجراء دراسة تحليلية لمحركات البحث مجال الدراسة. وقد قامت الباحثة بإجراء عمليات بحث في كل من محرك بحث حكيا وجوجل باستخدام ٤ مصطلحات بحث مختلفة بغرض التعرف على ما يميز محرك بحث حكيا من حيث خصائص استرجاع المعلومات وعرض النتائج.

مفهوم البحث الدلالي:

قبل ان نتناول التعريف بالبحث الدلالي semantic search ، يجدر بنا أن نشير إلى المقصود بالويب الدلالي، أو ما يعرف بالويب ذو الدلالة اللفظية، أو الويب ذو المعنى semantic web ، والمقصود بذلك هو اعتمادها على برمجيات يمكنها تعريف المقصود بالبيانات التي تقدمها الشبكة العنكبوتية (الويب)، وذلك بالاستعانة بما يعرف بخرائط المفاهيم ontology . ويمثل الويب الدلالي أحد فروع الذكاء الاصطناعي ويعد ثورة في عالم الويب حيث يسمح للمتصفح

أو البرامج الوكيلة بالبحث عن المعلومات، وبالتالي فإن عملية معالجة المعلومات تتم اعتماداً على الحاسبات الآلية بدلاً من البشر.

والبحث الدلالي هو أحد التوجهات الحديثة التي تعتمد على تطبيقات الويب الدلالي لاسترجاع المعلومات من النظام، وذلك اعتمادا على دلالة المصطلحات التي يرغب المستفيد الحصول على نتائج حولها، فبدلاً من أن يسترجع النظام نتائج اعتماداً على الترتيب الشائع للمواقع؛ فإن التركيز في البحث الدلالي يكون على تقديم نتائج تتفق مع معاني الكلمات. وبذلك فإن المستفيد عندما يوجه استفساراً إلى نظام الاسترجاع يتضمن كلمة أو عبارة فإن آلية البحث الدلالي تهدف إلى تقديم النتائج الأكثر صلة باستفساره، وذلك وفقاً لمعانى الكلمات التي يقوم المستفيد بالبحث عنها [١].

ويمكن ان نعرف البحث الدلالي على أنه عملية تستخدم لتحسين البحث المباشر عن طريق استخدام بيانات من شبكات دلالية لتزيل الغموض من الاستفسار ومن نص الويب من أجل الحصول على نتائج ذات صلة أكبر بالاستفسار[٢].

متطلبات الويب الدلالي:

يتطلب تطبيق الويب الدلالي استخدام بعض التقنيات التي تعد بمثابة أدوات تساعد على تحقيق الدور تقوم به الويب الدلالي، وفيما يأتي نوضح أربع متطلبات لازمة لعمل الويب الدلالي، وتتمثل في الآتي:

١) لغة الترميز القابلة للتوسع
٢) إطار وصف المصادر
٣) خرائط المفاهيم (الأنطولوجيا)
٤) لغة أنطولوجيا الويب

أولاً: لغة الترميز القابلة للتوسع Extensible Markup Language (XML) :

تتفرع من اللغة المعيارية العامة لتهيئة النصوص SGML : Standard Generalized Markup Language ، وتستخدم XML لوصف بنية ومحتوى الوثائق الإلكترونية المتاحة على الويب، وتتميز بالمرونة الشديدة، حيث تسمح لمستخدمها بإعداد الرموز التي يحتاج إلها، وإنشاء العناصر الملائمة لوصف الوثائق وبناء محتواها، فهي لا تضم عناصر

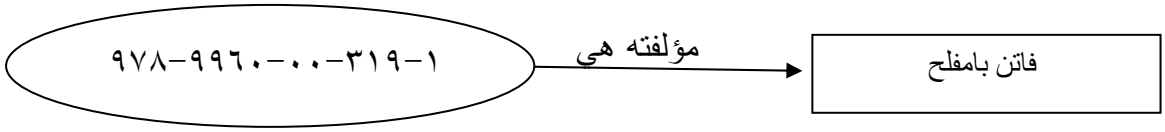
ثابتة محددة مسبقاً كما هو الحال في لغة ترميز النص الفائق، ولكنها تسمح بابتكار عناصر الترميز الملائمة للعمل وتسميتها وفقاً لاختيار القائمين بترميز الوثيقة[٣].

ثانياً: إطار وصف المصادر Resource Description Framework (RDF) :

يحتاج الويب الدلالي إلى استخدام نظام ميتاديتا لتنسيق عملية تبادل البيانات، ويعد إطار وصف المصادر RDF أحد أبرز الأدوات المستخدمة مع تطبيقات الويب الدلالي. وقد تم تطوير RDF من قبل اتحاد الشبكة العنكبوتية world wide web consortium (W3C) ، لتعمل مع لغة XML ، وتكون بمثابة إطار عمل يسمح باستخدام أكثر من معيار مختلف من معايير ميتاديتا بشكل متكامل مع بعضهم البعض [٤]. وباستخدام RFD يتم ترميز عناصر المعلومات المكتوبة بلغة XML ، بتجزئتها إلى بيانات تمثل أجزاء صغيرة يتم حفظها وفقاً لقواعد محددة توضح دلالاتها أو معانيها[٥].

ويوضح تيم بيرنرز لي Tim Berners- Lee مطور الويب الدلالي في إطار وصفه لكل من لغة XML و إطار وصف المصادر، أن هاتين التقنيتين مهمتين لتطوير الويب الدلالي، حيث تسمح لغة الترميز القابلة للتوسع XML بهيكلة الوثائق وإضافة وسيمات لها، ولكنها لا توضح شيء عما تعنيه تلك الهيكلة، أما المعنى فيتم التعبير عنه بواسطة إطار وصف المصادر RDF[٦].

ويعمل إطار وصف المصادر اعتماداً على ثلاث عناصر هي: المورد ونوع الخاصية وقيمة الخاصية، فعلى سبيل المثال: كتاب المكتبات الرقمية ذو الترقيم المعياري الدولي ١-٣١٩-٠٠-٩٩٦٠-٩٧٨ ، مؤلفته فاتن بامفلح، وهنا المورد هو: كتاب المكتبات الرقمية الذي يحمل الترقيم المعياري ١-٣١٩-٠٠-٩٩٦٠-٩٧٨، والخاصية هي أن له مؤلف، وقيمة الخاصية هي اسم المؤلف فاتن بامفلح. ويعد الترقيم المعياري هنا بمثابة Universal Resource Identifier (URI) ، وقد يكون المعرف عنوان Uniform Resource Locator (URL).
ويمكن تمثيل العناصر الثلاثة في تعليمات RDF على النحو الآتي:



وتظهر هذه البيانات بالترميز على النحو الآتي:

```
<rdf:RDF>
 <rdf:Statement>
  <rdf:subject   rdf:resource="978-9960-00-319-1" />
  <rdf:predicate rdf:resource="onto;its author is" />
```

```
  <rdf:object    rdf:resource="Faten Bamofleh" />
 </rdf:Statement>
</rdf:RDF>
```

وبذلك فإن الخواص تقوم بوصف الموارد، بحث توجد لكل خاصية ميزة أو سمة أو علاقات محددة مع المورد، وكذلك فإن لها معنى خاص بها يمكن التعرف عليه من خلال اسم الخاصية والمعرف URI المرتبط بها[٧].

ثالثاً: خرائط المعرفة (الأنطولوجيا) Ontology :

هي أدوات لتمثيل المعرفة تحصر المصطلحات التي تعبر عن الموضوعات المعرفية والعلمية، وتنظمها ، موضحة العلاقات المختلفة التي تربط بينها؛ حيث توضح المصطلحات المترادفة، وذات الصلة ، والأعرض ، والأضيق، كما تحلل الصيغ المختلفة من المصطلحات؛ بما في ذلك : الأسماء والأفعال والصفات والظروف وغيرها...

وتعرف الأنطولوجيا أيضاً على أنها : وصف للمعارف في مجال معين بتقسيمها إلى فئات تمثل مفاهيم concepts ، ولكل مفهوم خصائص وسمات تعرف بالأدوار roles ، أو الخصائص properties . وهناك قيود لتلك الأدوار أو الأوجه.

ويصف البعض خريطة المفاهيم من حيث الشكل والمضمون على أنها عبارة عن مخطط ثنائي البعد، ترد فيه المفاهيم محددة بالمصطلحات التي تعبر عنها، ومنظمة في شكل هرمي متدرجة من العام إلى الخاص، فتأتي على رأس الهرم المفاهيم العامة ، وتليها في المستويات الهرمية المفاهيم الفرعية والأمثلة نزولاً إلى قاعدة الهرم. وتوضح خريطة المفاهيم العلاقات بين المفاهيم الواردة في المستويات المختلفة باستخدام أسهم تربط المصطلحات ببعضها ، مع كتابة كلمات رابطة على الأسهم لإيضاح طبيعة العلاقة بين المفاهيم، ومن أمثلة تلك الكلمات (يؤثر في، ينقسم، يتكون من، يتضمن ...) [٨].

وتحقق خرائط المفاهيم (الأنطولوجيا) عديد من أغراض الويب الدلالي والمتمثلة في الآتي:

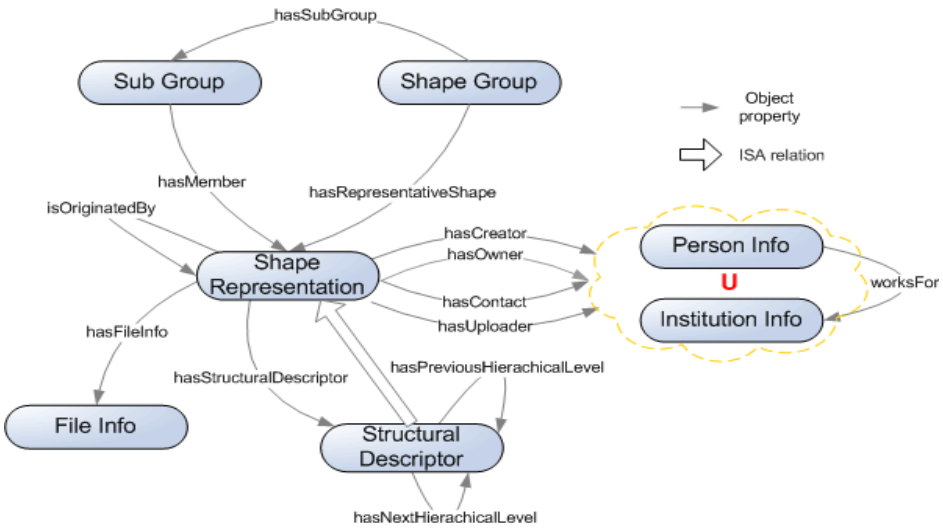
أ. تقلل من درجة الغموض الدلالي للمصطلحات إلى الحد الأدنى.

ب. تعزز إمكانية التشغيل التبادلي بين الأنظمة في ميادين المعرفة المختلفة.

ج. تستخدم لإنشاء البرامج الوكيلة الذكية التي تعمل على إنجاز أعمال محددة.

وتنقسم خرائط المعرفة إلى نوعين؛ أحدهما خاص بمجال محدد، وبالتالي يتضمن معاني المصطلحات وفقاً لذلك المجال، أما النوع الآخر فهو عام يمثل المعرفة الممتدة عبر المجالات المختلفة ويعبر عن المعاني عبر ذلك المدى العريض[٩]. ويتطلب تطبيق الويب الدلالي ربط كل

صفحة من صفحات الويب بخريطة مفاهيم توضح ما تتناوله الصفحة من مفاهيم؛ الأمر الذي يتيح للمستفيد عند استرجاع مفهوم معين من خلال محركات البحث، أن يصل الى نتائج لصفحات قد لا يكون ذكر فيها المصطلح في حد ذاته، ولكن خريطة المفاهيم الخاصة بالصفحة تشير إلى أن هناك ارتباط بين محتوى الصفحة والمفاهيم التي تغطيها وبين ما يستفسر عنه المستفيد. وشكل (١) يوضح نموذج لخريطة المعرفة.



شكل (١)

نموذج لخريطة المعرفة [١٠]

رابعاً: لغة أنطولوجيا الويب Ontology web Language (OWL) :

هي بمثابة مخطط للعلاقات يعمل على تسهيل وصف وتمثيل خرائط المفاهيم والمصطلحات والعلاقات التي تربط بينها ضمن مجال محدد. وتعد هذه اللغة أحدث من إطار وصف المصادر RDF –الذي يمكن أن يستخدم أيضاً لتمثيل الخرائط– فهي تقدم مميزات أوسع منها في تمثيل المفاهيم، حيث تركز OWL على (ربط المعلومات مع بعضها، وتأسيس العلاقات بين المعلومات والمفاهيم التي تجمعها علاقة ما، وذلك اعتماداً على قوانين مقتبسة من علوم فلسفية، وهذه العلاقات والروابط التي تؤسس بين المعلومات تستفيد منها البرمجيات المختلفة في فهم وبالتالي تحليل ومعالجة المعلومات طبقاً للعلاقات التي تربطها) [١١].

محركات البحث الدلالي:

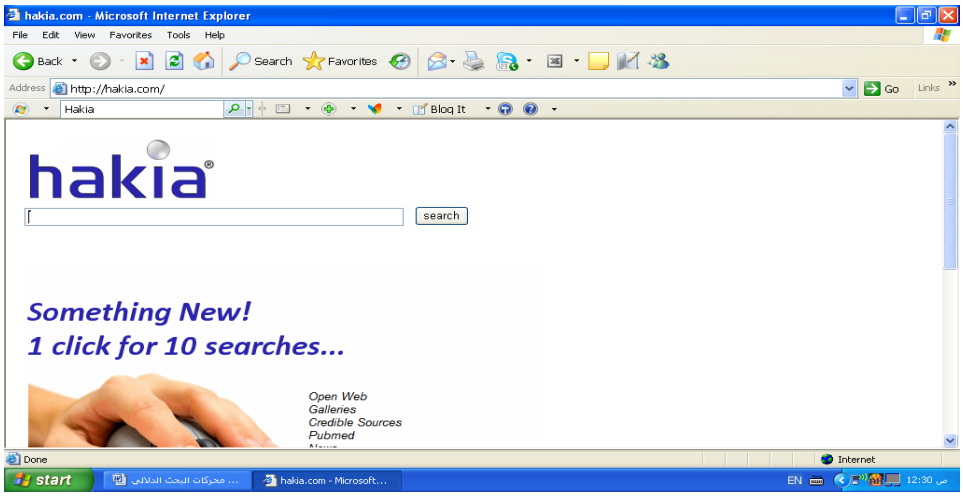
تضاهي محركات البحث بين استفسار المستفيد وبين بيانات الوثائق التي يضمها النظام، وتسترجع النتائج المطابقة للاستفسار. وفي حين تعتمد محركات البحث التقليدية مثل جوجل

والتفيستا وياهو على تطابق المصطلحات وتماثلها بين الاستفسار والنص، نجد أن محركات البحث الدلالي تضع في الاعتبار معاني ودلالات المصطلحات وليس تطابقها فقط، وبذلك فإن نتائج الأخيرة تكون ذات صلة أكبر بالاستفسار. وهناك عديد من محركات البحث الدلالي المتاحة من خلال شبكة الإنترنت، ومن بينها الآتي:

١. حكيا Hakia :

يقدم النتائج مرتبة باستخدام لوغاريتمات تمنح درجة scores للنتائج بناء على درجة صلتها ومطابقتها للمفهوم الوارد في الاستفسار (شكل ٢). وتعتمد حكيا على ثلاث تقنيات رئيسة تتمثل في الآتي:

- Ontosem : الأنطولوجيا أو مستودع المعاني sense repository ، يضم علاقات المفاهيم، وهو بمثابة قاعدة بيانات لغوية تضم الكلمات مرتبة في فئات وفقاً لمعانيها.
- QDEX : تقنية تكشيف الاستفسار، وتستخدمه حكيا كبديل للكشاف المقلوب المستخدم في معظم محركات البحث لحفظ محتويات الويب. ويعمل QDEX على اقتباس كل العلاقات المحتملة للاستفسار بالمحتوى.
- Semantic Rank algorithm : لوغاريثمات الترتيب الدلالي، وترتب المحتوى على أساس تحليل العبارات بشكل أكبر، وتضع في الاعتبار أيضاً الموثوقية وعمر المحتوى بغرض تحديد الصلة .



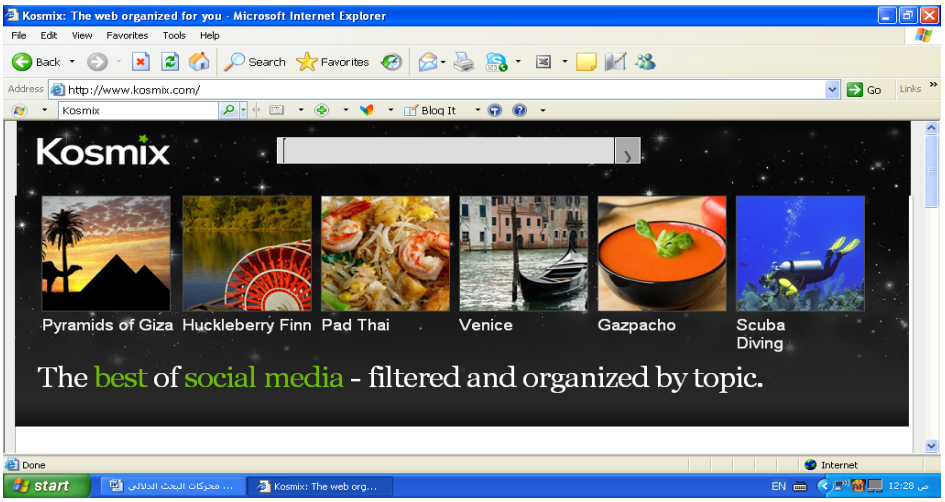
شكل (٢)

محرك حكيا

٢. كوزمكس Kosmix :

يقدم محرك البحث للمستفيد لوحة قياس للمحتوى، بالإضافة إلى فئات المفاهيم، ولا يقتصر على عرض إجابة على استفسار المستفيد وتقديم عناوين URL فقط، ولكنه

يقدم معلومات حول الموضوع أيضاً، حيث يقدم مزيج من الروابط links ، ومقاطع الفيديو، وال tweets للمستفيد. وشكل ٣ يوضح واجهة محرك البحث.

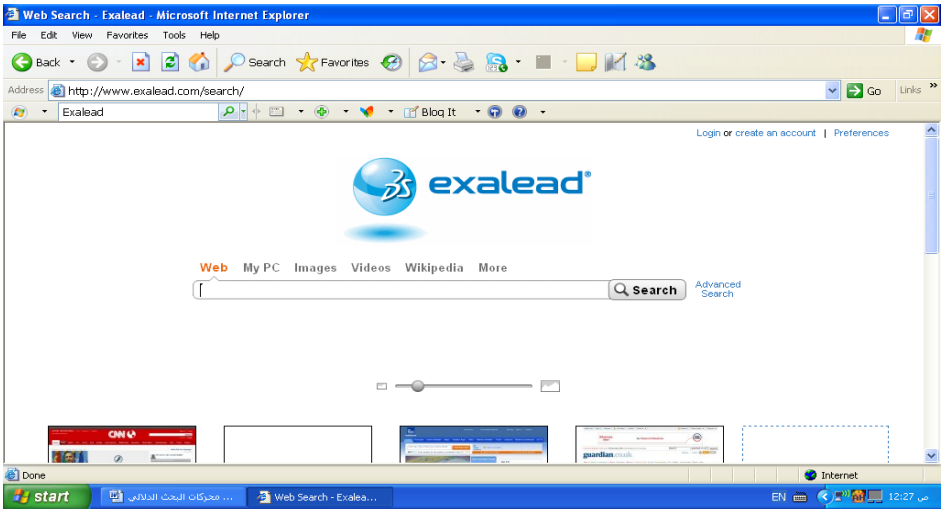


شكل (٣)

محرك كوزمكس

٣. إكساليد Exalead :

هو محرك بحث للصور (شكل ٤)، يسمح بتضييق البحث اعتماداً على حجم الصور ولونها ومحتواها، وبذلك فإنه يساعد على حل المشكلات المتعلقة بالبحث في محتوى الصور، خصوصاً في الحالات التي لا يكون فيها تحليل الروابط link analysis كافٍ لتقديم ما يبحث عنه المستفيد بدقة كافية.



شكل (٤)

محرك أكساليد

٤. سووقل Swoogle:

محرك بحث خاص بالبحث في الويب الدلالي semantic web search (شكل٥)، حيث يعمل على تكشيف الوثائق التي تم تطويرها وفقاً لمفاهيم ومعايير الويب الدلالي، ومن ذلك على سبيل المثال تلك التي تستخدم RDF format [١٢].

شكل (٥)

محرك سووقل

الفرق بين محركات البحث الدلالي والتقليدي:

تختلف محركات البحث الدلالي عن نظيرتها التقليدية من جوانب عدة، منها ما يتعلق بتقنيات تطوير تلك المحركات، وكذلك آلية عملها، والنتائج التي تقدمها للمستفيد...الخ. وفيما يأتي نوضح ما يميز محركات البحث الدلالي عن نظيرتها التقليدية.

أولاً: تعمل محركات البحث الدلالي على ربط العلاقات بين الوثائق والأشخاص والأماكن والأحداث؛ فعلى سبيل المثال: عند استرجاع معلومات عن جنيفر لوبيز؛ فإن نتيجة البحث ستضم معلومات عنها، والبوماتها الغنائية، وجدول بمواعيد حفلاتها، والشركة التي تنتج لها أعمالها ، وأفلامها ...الخ. وبذلك فإن نتائج البحث التقليدي تأخذ شكل قائمة بالوثائق وصفحات الويب، وهذا يعني أن نتائج البحث الدلالي لا تقتصر على قائمة بالوثائق أو صفحات الويب فقط، ولكنها تتجاوز ذلك إلى تقديم البيانات ذات الصلة[١٣].

ثانياً: تستخدم محركات البحث الدلالي تقنيات معقدة بغرض استرجاع نتائج بحث مطابقة أو ذات صلة كبيرة باستفسار المستفيد، وتعد خرائط المفاهيم (الأنطولوجيا) من بين التقنيات التي تستخدمها تلك المحركات. ويتم تخزين المعلومات الداخلية في محركات البحث الدلالي في

شبكة دلالات باستخدام صيغ RDF و XML ، وتضم شبكة الدلالات نقاط nodes يعرف كل منها باسم synsets ، وتتضمن النقاط مترادفات للمعاني[١٤].

ثالثاً: لا يتناسب البحث الدلالي مع طريقة البحث الملاحي navigational search التي يحاول فيها المستفيد عادة استرجاع وثيقة معينة وبالتالي فإنه يستخدم كلمات متفرقة بدون معنى لضمان ورودها ضمن الوثيقة التي يتم استرجاعها، وفي المقابل فإن البحث الدلالي يلائم البحث البحثي research search الذي يستخدم فيه المستفيد في عملية الاسترجاع كلمة ذات معنى للوصول إلى نتائج حولها، وقد تكون تلك الكلمة اسم شخص أو مكان[١٥].

رابعاً: تعمل محركات البحث الدلالي على التعرف على معاني الكلمات والاسترجاع على أساس تلك المعاني، وهي بذلك تختلف عن محركات البحث التقليدية التي تبحث عن الكلمة او الكلمات المفتاحية وما يضاهيها في صفحات الويب، وبذلك فإن محركات البحث الدلالي تذهب إلى أبعد من استرجاع الكلمات المفتاحية وبيانات الوصف، حيث تعتمد على المحتوى والمضمون والعلاقات، وهذا من شأنه أن يجعل النتائج التي يسترجعها محرك البحث الدلالي ذات صلة أكثر باستفسار المستفيد؛ مما يوفر عليه الوقت المستغرق في قراءة صفحات غير ذات صلة قد يسترجعها محرك البحث التقليدي[١٦].

خامساً: تعمل محركات البحث الدلالي على إزالة الغموض من مصطلحات البحث، وذلك عن طريق تحليل المصطلحات الواردة في عبارة البحث بغرض الوصول إلى المعنى الأقرب لاستفسار المستفيد؛ فعلى سبيل المثال : كلمة bark قد تعني نباح الكلب أو لحاء الشجر ، وبذلك ففي حالة البحث عنها فإن محرك البحث الدلالي يعمل على تحليل السياق الذي جاءت فيه الكلمة ضمن الاستفسار، حيث يضع في الاعتبار معاني الكلمات الأخرى التي تظهر في الاستفسار وفي النص.

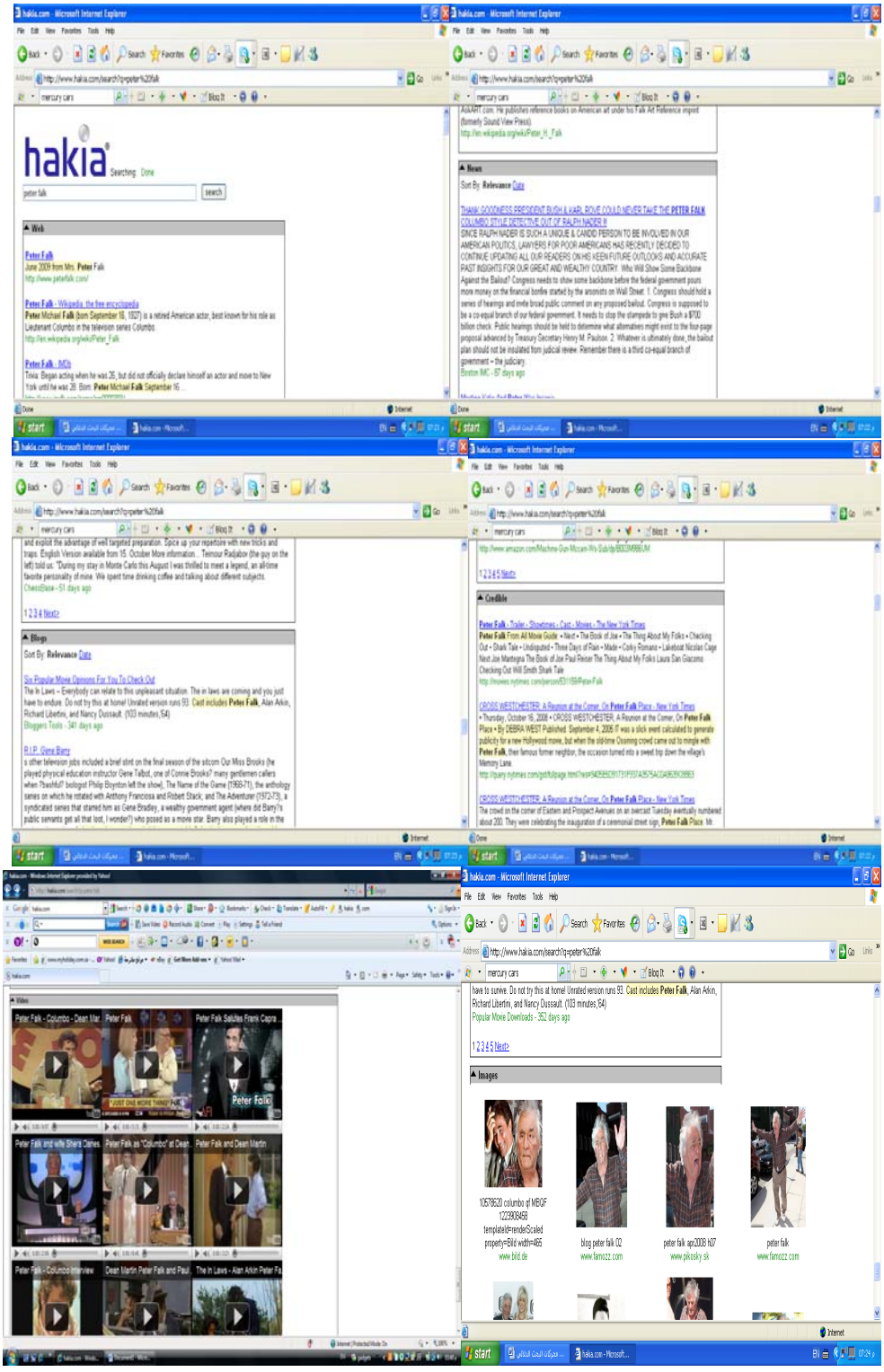
استرجاع المعلومات باستخدام محرك حكيا:

ذكر تيم بيرنرز لي مطور الويب الدلالي في إطار إجابته على تساؤل عن الحاجة للويب الدلالي في ظل وجود جوجل بإمكانياته الهائلة في الرد على استفسارات المستفيدين؛ أشار إلى أن جوجل تعد أداة رائعة للبشر، في حين أنها لا تخدم الآلة، والمقصود بذلك أنها من مواقع الويب المفهومة للبشر، أما بالنسبة للآلة فهي عبارة عن صفحات ممثلة باللغة الثنائية (الصفر

والواحد) دون أن يكون لها أي معنى، أما الويب الدلالي فيجعل الآلة تفهم محتويات صفحات الويب وما تعنيه الروابط في الصفحة[١٧].

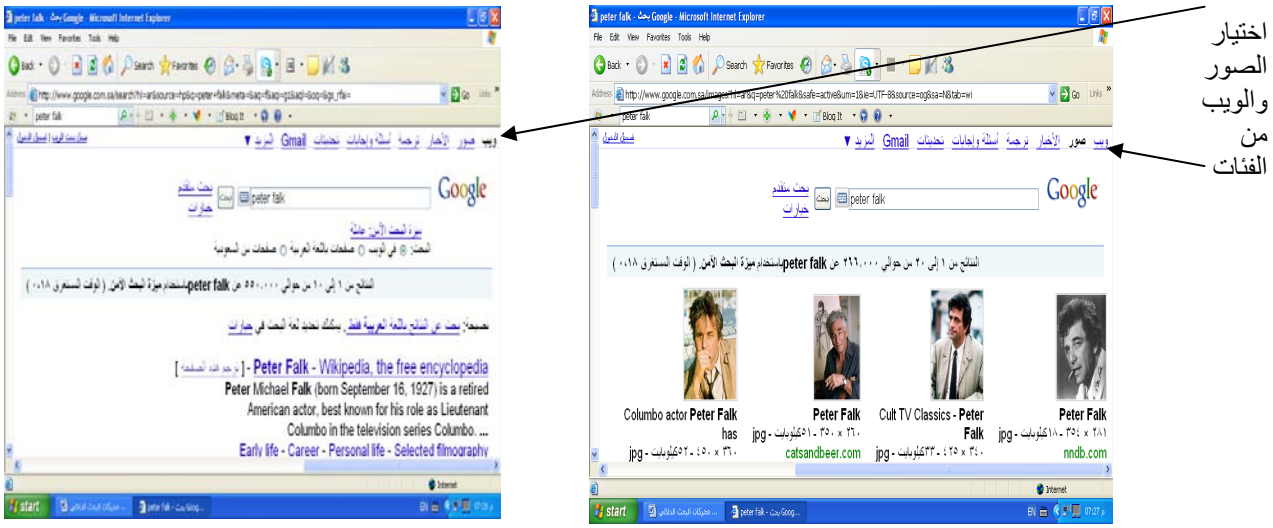
في محاولة للتعرف على نمط الاسترجاع في محركات البحث الدلالي بطريقة تطبيقية عملية، قامت الباحثة بالمقارنة بين استرجاع المعلومات في محرك بحث حكيا؛ الذي يمثل أحد محركات البحث الدلالي، ومحرك بحث جوجل الذي يمثل النمط التقليدي لمحركات البحث . واستخدمت الباحثة أربع مصطلحات بحث مختلفة هي: "Peter Falk" ، و " Library of Congress " و "Paris Hilton"، و "Mercury Car "، وقد حرصت الباحثة على تنوع المصطلحات لتشمل اسم شخصية معروفة هي الممثل بيتر فالك، ومؤسسة ثقافية هي مكتبة الكونجرس، واسم يحتمل أكثر من معنى حسب التباديل وهو باريس هيتلون، وكذلك سيارة ميركوري، والغرض من تنوع المصطلحات وتعددها هو حرص الباحثة على للتعرف على الخصائص المختلفة والمميزات التي يمنحها محرك حكيا مقارنة بجوجل. وتبين من عملية الاسترجاع ما يأتي:

١. جاء عرض نتائج البحث في حكيا بشكل متكامل، حيث جاءت نتيجة البحث عن مصطلح peter falk على سبيل المثال؛ مقسمة الى فئات وفقاً للآتي: الويب، والمواد الموثوقة credible ، والأخبار والمدونات والصور، والفيديو..، وقد يضاف إليها فئات أخرى في نتائج البحث حول موضوعات أخرى، ومن ذلك المعارض Gallery، والويكيبديا wikipedia عند البحث عن مصطلح باريس هيلتون على سبيل المثال. وقد تظهر فئات أخرى مثل مواقع الشبكات الاجتماعية Twitter في حالة توافر ما يرتبط منها بمصطلح البحث. أما في محرك جوجل فإن نتيجة البحث حول المصطلح الواحد تتطلب من المستفيد اختيار الفئة التي يرغب البحث فيها، ومن ذلك الويب أو الصور أو الأخبار أو أسئلة وإجابات حول الموضوع ...، وشكل (٦ و٧) يوضح نتيجة البحث عن مصطلح بيتر فالك في كل من حكيا وجوجل.



شكل (٦)

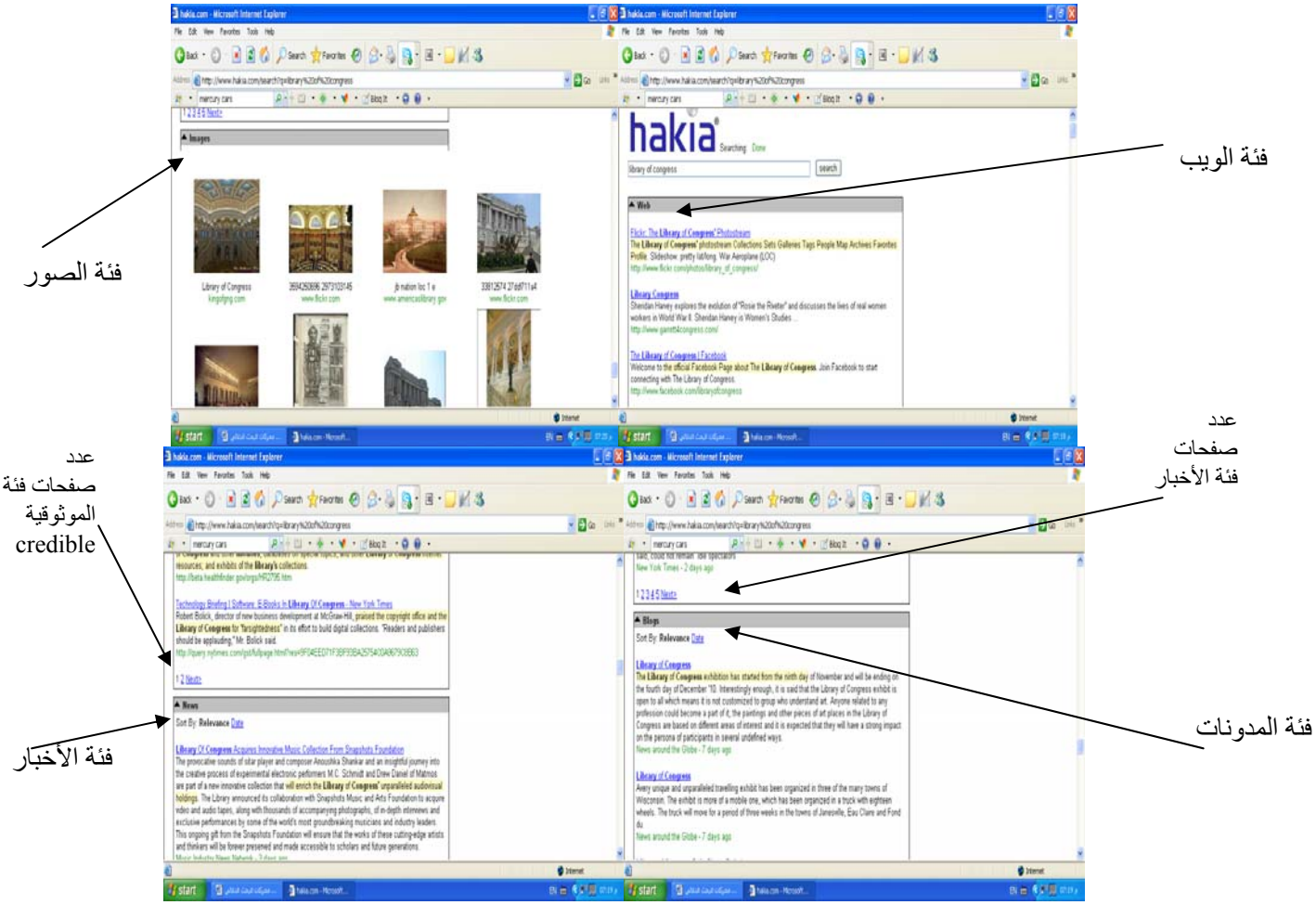
نتيجة البحث عن peter falk في حكيا تظهر الفئات المختلفة متتابعة في النتيجة نفسها



شكل (٧)

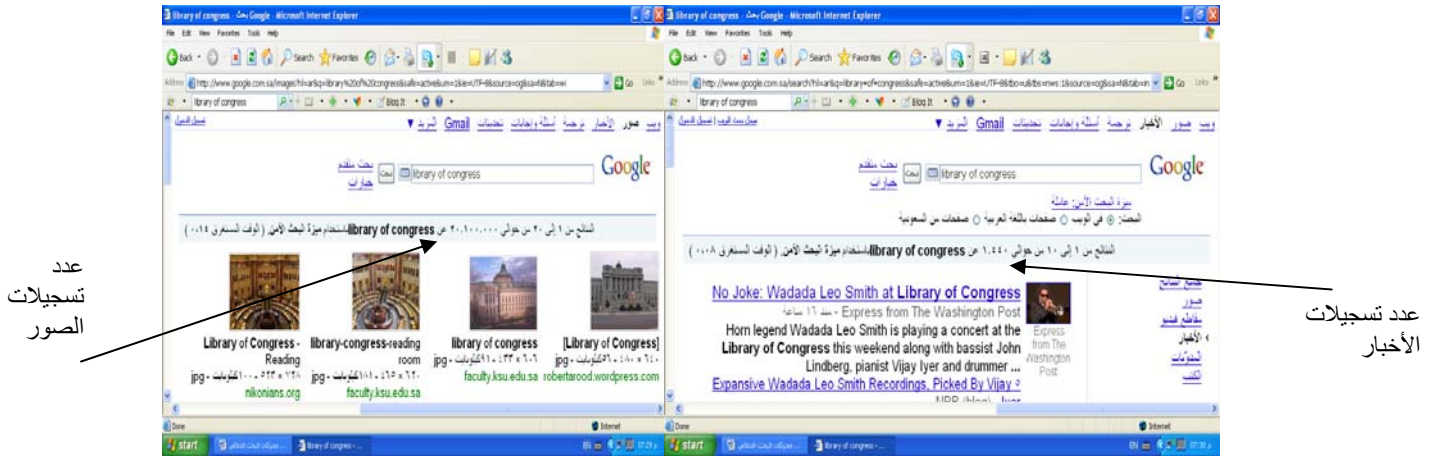
نتيجة البحث عن peter falk في جوجل وتظهر نتيجة الويب مستقلة عن الصور

٢. أن نتائج البحث في محرك حكيا جاء أقل من نتائج محرك جوجل من حيث العدد، ولعل السبب في ذلك يرجع إلى ضخامة قاعدة بيانات جوجل، وكذلك إلى تركيز نتائج حكيا أكثر حول موضوع الاستفسار وما يتعلق به، سواء للمواد النصية أم المصورة أم المسجلة فيديو... فعلى سبيل المثال بلغت نتيجة البحث عن مكتبة الكونجرس في جوجل ١٧.٣٠٠.٠٠٠ تسجيلة لمواقع الويب فقط، بالإضافة إلى ٢٠.١٠٠.٠٠٠ صورة، و١.٤٤٠ تسجيلة أخبار، في مقابل أقل من ١٥٠ تسجيلة فقط في حكيا تشمل مواقع الويب والأخبار والمدونات والصور وغير ذلك، حيث لم يتجاوز عدد مواقع الويب ٥٠ تسجيلة، وعدد الصور ١٠، والأخبار ٥٠ تسجيلة... وشكل (٨، ٩) يوضح نتائج البحث في كل من حكيا وجوجل ، ويلاحظ أن محرك حكيا لا يعرض عدد التسجيلات الناتجة عن عملية البحث، ولكنه يورد ١٠ تسجيلات في كل صفحة لكل فئة مع تحديد عدد صفحات النتائج أسفل كل فئة.



شكل (٨)

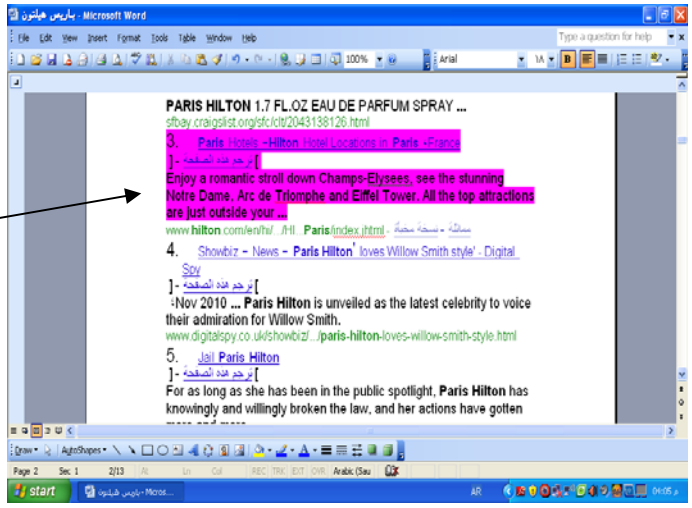
جانب من نتيجة البحث عن مكتبة الكونجرس في حكيا



شكل (٩)

جانب من نتيجة البحث عن مكتبة الكونجرس في جوجل

٣. عند استرجاع معلومات باستخدام مصطلحات تحتمل أكثر من معنى؛ فإن دقة محرك حكيا تتضح بشكل أكثر، فعلى سبيل المثال عند استرجاع معلومات حول باريس هيلتون، فإن الأمر يحتمل وفقاً لتباديل مصطلح البحث أن يكون المطلوب هو معلومات عن الممثلة باريس هيلتون مالكة فنادق هيلتون العالمية، كما يحتمل أيضاً أن يكون الاستفسار حول معلومات عن فندق هيلتون في مدينة باريس. وباسترجاع معلومات باستخدام مصطلح Paris Hilton ظهر الفرق بين استخدام محرك البحث الدلالي وبين نظيره التقليدي ففي محرك حكيا اقتصرت النتائج التي استرجعها النظام حول الفنانة باريس هيلتون، حيث تمت مراعاة السياق الذي جاءت فيه الكلمتين وترتيبهما ضمن مصطلح البحث، وجاءت النتيجة شاملة لمواقع الويب التي تناولت الفنانة، وأخبارها، وصورها، ومقاطع فيديو حولها...، وفي المقابل فإن البحث في جوجل جاء بنتائج تتعلق بالفنانة باريس هيلتون، وأخرى حول فندق هيلتون. وشكل ١٠ و ١١ يوضح نتائج البحث في جوجل. وشكل ١٢ و١٣ يوضح نتائج البحث في حكيا.

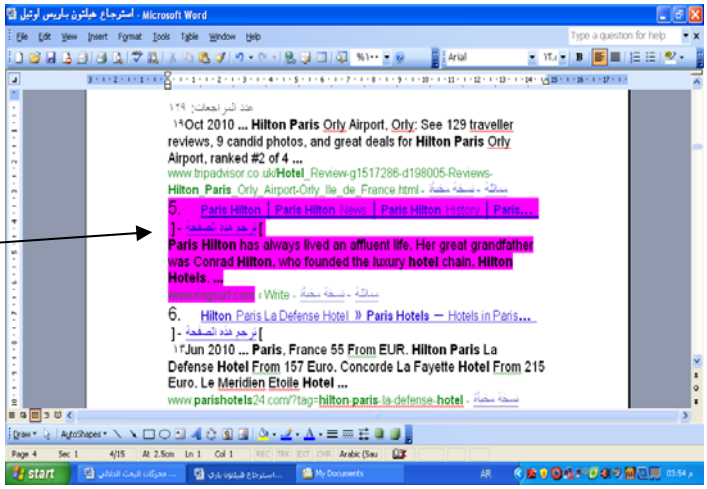


تسجيلة عن فندق هيلتون تظهر بين النتائج

شكل (١٠)

نتيجة البحث في جوجل عن باريس هيلتون

وتظهر فيها نتائج حول فندق هيلتون في باريس إلى جانب النتائج حول الممثلة باريس هيلتون

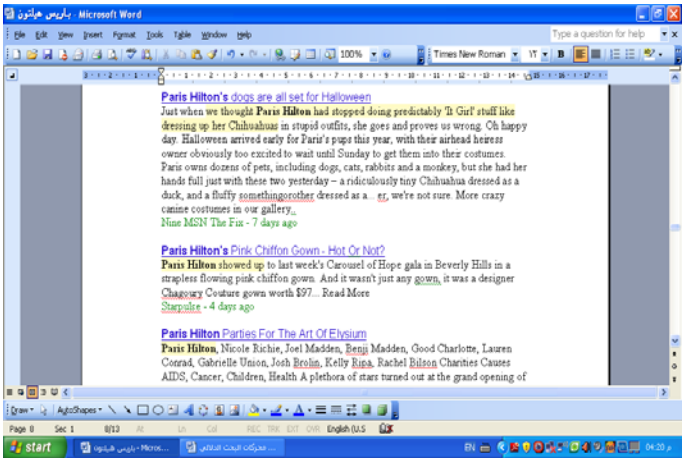


تسجيلة عن الممثلة باريس هيلتون تظهر بين النتائج

شكل (١١)

نتيجة البحث في جوجل عن مصطلح هيلتون باريس

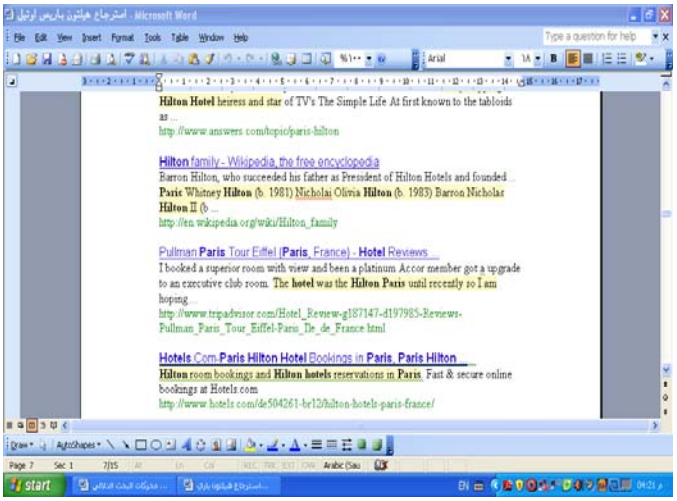
وتظهر فيها نتائج حول الممثلة باريس هيلتون في باريس إلى جانب النتائج حول فندق هيلتون في باريس



شكل (١٢)

نتيجة البحث في حكيا عن باريس هيلتون

تقتصر فيه النتيجة على تسجيلات حول الممثلة باريس هيلتون

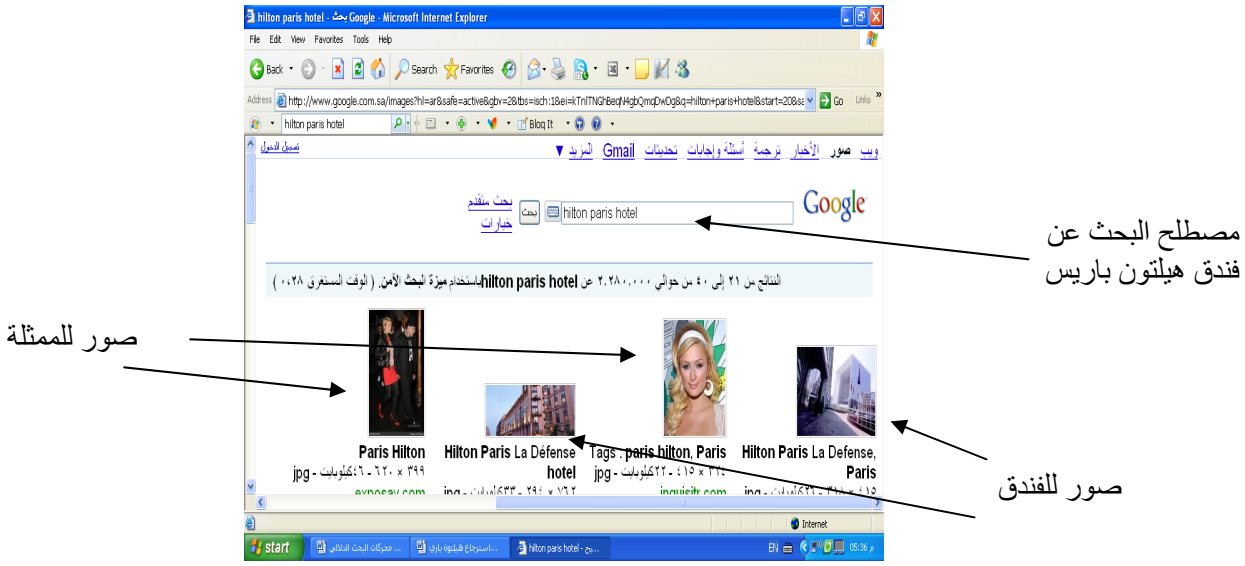


شكل (١٣)

نتيجة البحث في حكيا هيلتون باريس

تقتصر فيه النتيجة على تسجيلات حول فندق هيلتون في باريس

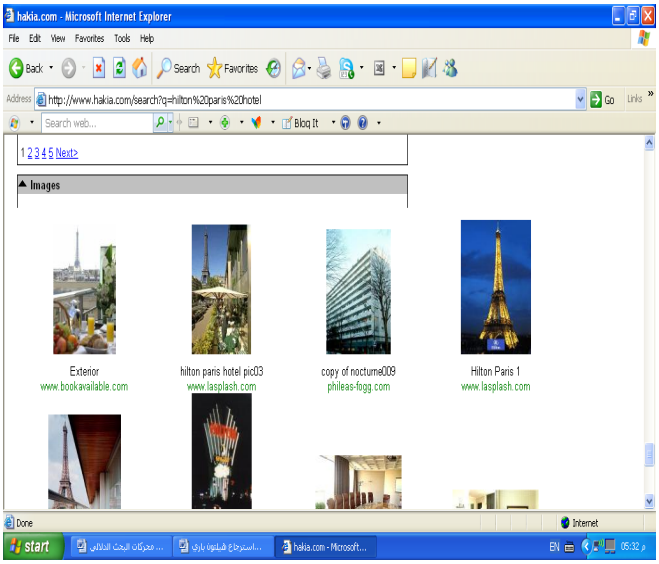
ولم يقتصر الأمر على استرجاع النصوص فقط، ولكنه ظهر جلياً في نتائج الصور والفيديو عند استخدام مصطلح Hilton Paris hotel ، حيث أن نتائج البحث في جوجل جاءت بصور للممثلة باريس هيلتون، بالإضافة إلى صور الفندق (شكل ١٤)، في حين اقتصرت نتيجة البحث في حكيا على صور حول الفندق فقط (شكل ١٥)، حيث وضع في الاعتبار السياق الذي جاءت فيه المصطلحات، وورود كلمة فندق ضمن مصطلح البحث.



شكل (١٤)

نتيجة البحث في صور جوجل عن مصطلح فندق هيلتون باريس

تظهر في النتيجة صور للممثلة باريس هيلتون ضمن صور الفندق



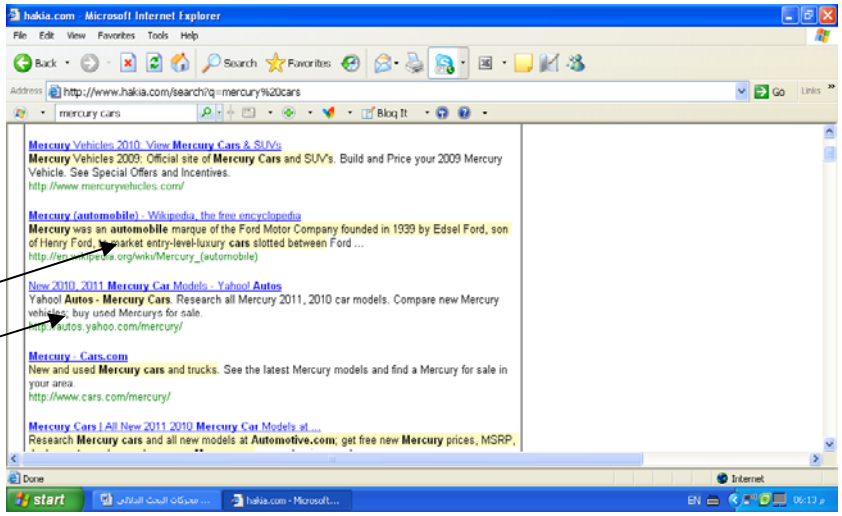
شكل (١٥)

نتيجة الصور في حكيا بالبحث عن مصطلح فندق هيلتون باريس

تقتصر النتيجة على صور الفندق

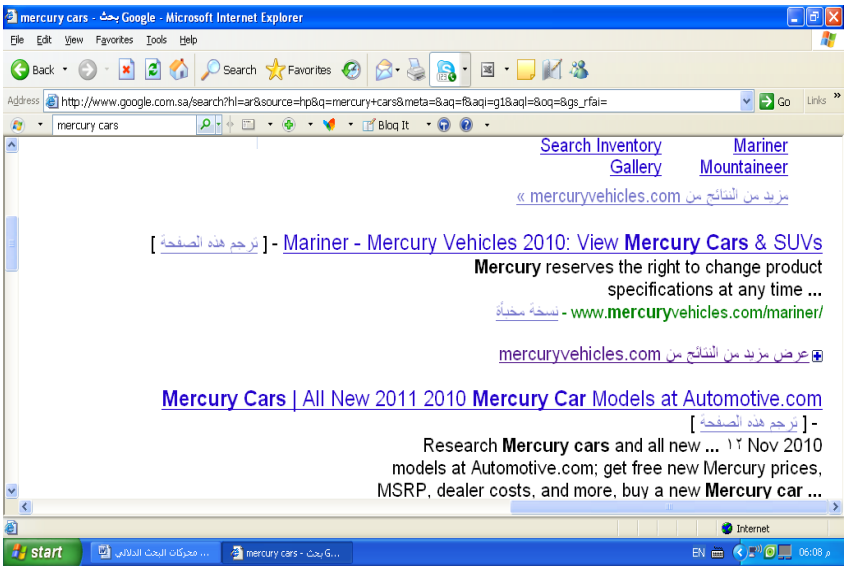
٤. عند استرجاع معلومات حول سيارات ميركوري Mercury cars من محرك بحث حكيا فإن عملية البحث استرجعت نتائج تضمنت مصطلح mercury cars ، بالإضافة إلى نتائج تضم كلمة اوتوموبيل automobile أو كلمة autos بدلاً من cars ، وذلك على اعتبار أن الثلاث مصطلحات مترادفة، بل إن محرك بحث حكيا قام بتظليل الكلمات الثلاثة عند ورودها ضمن أي تسجيلة، على اعتبار أنها تمثل المصطلح الذي استفسر المستفيد عنه. والأكثر من ذلك أن نتيجة البحث تضمنت تسجيلات لم يرد فيها أي من المصطلحات السابق ذكرها، ولكن ورد فيها اسم الشركة المصنعة لسيارات ميركوري وهي Ford ، أو اسم موديل من الميركوري مثل جراند مركيز Grand Marquis ، وهذا يدل على أن اعتماد محرك بحث حكيا على الأنطولوجيا جعله يتعامل مع دلالة المصطلح المستخدم من قبل المستفيد وهو في هذه الحالة mercury cars ، ليسترجع المصطلحات الأخرى التي تحمل نفس الدلالة، أما نتيجة البحث في جوجل فقد تضمنت فقط المصطلحات الواردة ضمن الاستفسار وهي mercury cars ، مع مراعاة ورود كلا الكلمتين في جميع نتائج البحث، وذلك على اعتبار أن محرك جوجل يقوم تلقائياً باسترجاع جميع المصطلحات الواردة في الاستفسار، إلاّ إذا اختار المستفيد من وواجهة البحث المتقدم أن يتم البحث عن أي من المصطلحات الواردة في الاستفسار. وشكل ١٦ و١٧

يوضح نتيجة البحث عن مصطلح سيارات ميركوري في كل مـــن محـــرك حكيـــا وجوجل.



شكل (١٦)

نتيجة البحث عن مصطلح سيارات ميركوري في حكيا



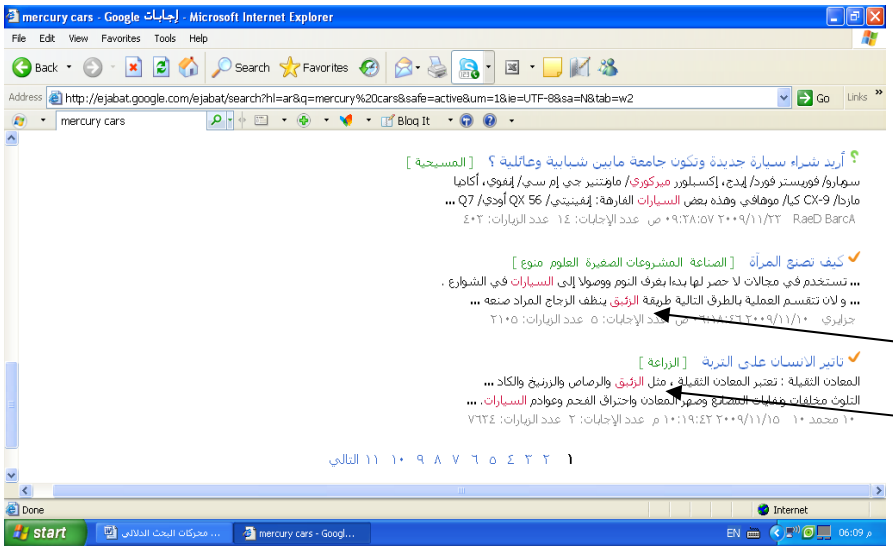
شكل (١٧)

نتيجة البحث عن مصطلح سيارات ميركوري في جوجل

وقد اقتصرت النتيجة على المصطلحات نفسها فقط

٥. لم تتضمن نتيجة البحث عن سيارات ميركوري في محرك حكيا أي تسجيلات تحمل دلالات أخرى لمصطلح ميركوري كإسم شخص أو لقب او الزئبق ، ولكـــن جـــاءت

النتائج جميعها مرتبطة تماماً بالاستفسار. أما في جوجل فقد جاء المصطلح مترجما بمعنى الزئبق ضمن نتائج الأسئلة والإجابات. وشكل ١٨ يوضح ذلك.



شكل (١٨)

نتيجة البحث عن مصطلح سيارات ميركوري في أسئلة وإجابات جوجل

تظهر فيها كلمة الزئبق بدلاً من ميركوري

٦. حاولت الباحثة إجراء بحث باللغة العربية في محرك حكيا، وذلك لمقارنة إمكانيات البحث باللغة العربية في النظام، ولكن تبين أن محرك البحث لا يتعامل مع اللغة العربية، فبالبحث عن مصطلح "مكتبة الكونجرس"، لم تسترجع عملية البحث أي نتيجة، وباستخدام مصطلح "عاصفة الصحراء" جاءت النتيجة بعدد قليل من التسجيلات في الويب فقط، ولكنها مكتوبة في شكل أكواد غير مقروءة، وحتى عند الضغط على روابط تسجيلات النتائج، فإنها كانت تحليل إلى صفحات غير قابلة للقراءة حيث جميعها تحتوي على رموز غير مقروءة.

النتائج والتوصيات:

توصلت الدراسة إلى النتائج الآتية:

١. يقدم البحث الدلالي نتائج أكثر صلة وارتباطاً باستفسارات المستفيدين، وذلك لاعتماده على دلالات مصطلحات البحث بدلاً من الاعتماد على الترتيب الشائع للمواقع.

٢. يعتمد البحث الدلالي في عمله على تطبيقات الويب الدلالي، ومن أبرزها لغة الترميز القابلة للتوسع XML و إطار وصف المصادر RDF، والأنطولوجيا.
٣. تختلف محركات البحث الدلالي عن نظيرتها التقليدية في عدة جوانب من بينها: تقنيات تطوير المحركات، وآلية عملها، والنتائج التي تقدمها للمستفيد.
٤. لا يتناسب البحث الدلالي على طريقة البحث الملاحي navigational search، ولكنه يتلاءم مع البحث البحثي research search .
٥. تقدم محركات البحث الدلالي نتائج أقل عدداً، ولكنها أكثر ارتباطاً بمصطلحات البحث، خصوصاً في ظل استخدام مصطلحات تحمل أكثر من معنى.

وفي ضوء تلك النتائج، توصي الدراسة بالآتي:

١) إجراء المزيد من الدراسات حول تطبيقات الويب الدلالي عموماً، والبحث الدلالي على نحو الخصوص.
٢) تطوير محركات بحث دلالي لاسترجاع المعلومات باللغة العربية.
٣) إجراء دراسة مقارنة لتقييم أداء عينة من محركات البحث الدلالي، وقدرتها على استرجاع المعلومات باللغة العربية.

قائمة مصادر البحث

[1] الويب الدلالي semantic web .- ع٨ .- مجلة كلية الهندسة المعلوماتية .- متاح على: http://www.alepposoft.net/itmag/2009/08 (٢٠٠٩/٣/١١)

[2] Semantic Web . Wikipedia , the free encyclopedia .- available at: http://en.wikipedia.org/wiki/Semantic_Web (20/ 10/ 2010)

[3] بامفلح ، فاتن سعيد . المكتبات الرقمية بين التخطيط والتنفيذ .- الرياض: مكتبة الملك فهد الوطنية، ٢٠٠٨م .-ص٢١١-٢١٢

[4] هودج ، جيل . فهم ما وراء البيانات (الميتاديتا) = understanding metadata / ترجمة جبريل بن حسن العريشي .- الرياض: مكتبة الملك فهد الوطنية، ٢٠٠٥م .- ص٢٣

[5] RDF .- available at: http://semanticweb.org/wiki/RDF (20/ 10/ 2010)

[6] Berners- Lee , Tim & Hendler , James & Lassila , Ora . The Semantic Web .- Scientific American (may 17/2001).- available at: http://www.scientificamerican.com/article.cfm?id=the-semantic-web

[7] هارولد، إيليوت رستي . إكس إم إل XML Bible / ترجمة خالد العامري .- القاهرة: دار الفاروق للنشر والتوزيع ، ٢٠٠٠م .- ص٧٩٦-٧٩٧

[8] Noy , Natalya F. & McGuinness , Deborah L. Ontology Development 101 : a guide to creating your first ontology .- available at: http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html (10/3/2010)

[9] عبدالهادي ، محمد فتحي . الاتجاهات الحديثة في التحليل الموضوعي للمعلومات وموقف قطاع المعلومات العربي منها .- في أعمال المؤتمر العشرين للاتحاد العربي للمكتبات والمعلومات (أعلم) : نحو جيل جديد من نظم المعلومات والمتخصصين : رؤية مستقبلية – الدار البيضاء ٩-١٠ ديسمبر ٢٠٠٩م مج١/ إشراف حسن عواد السريحي؛ تحرير فاتن سعيد بامفلح، الرياض: مكتبة الملك عبدالعزيز العامة،٢٠٠٩م .- ص٦٨٤-٦٨٦ .

[10] Common shape Ontology.- available at: http://dsw.aimatshape.net/tutorials/shape-ontology.gif

[11] الصقعوب، عبدالسلام بن عبدالعزيز . الويب الدلالي ثورة المعلومات الحقيقية .- جريدة الرياض: متاح على: http://www.alriyadh.com/2008/11/06/article385921.html (11/3/2010)

[12] Radhakrishnan , Arun . 9 Semantic Search engines that will change the world of search .- April 13th 2009 .- available at: http://www.searchenginejournal.com/semantic-search-engines/9832/ (15/6/2010)

[13] Guha , R. & McCool , Rob & Miller , Eric . Semantic Search .- available at: http://www2003.org/cdrom/papers/refereed/p779/ess.html

[14] What's Semantic Search ? Can Hakia Compete with Google? Available at: http://www.bulentongun.com/what-is-semantic-search-hakia-vs-google/ (11/3/2010)

[15] Guha , R. & McCool , Rob & Miller , Eric . op. cite

[16] What's Semantic Search ? Can Hakia Compete with Google?.- op.cite

[17] عبدالحميد، رجب . تقنيات الويب الدلالي للمكتبات الرقمية .- Cybrarian Journal .- ع١٤ (سبتمبر ٢٠٠٧م) .- متاح على: http://journal.cybrarians.info/index.php?option=com_content&view=article&id=88:2010-06-28-13-02-28&catid=43:2010-06-28-11-24-14&Itemid=56 (٢٠١٠/٩/١٥)