

نظم استرجاع معلومات الويب المبنية على الإدراك

دراسة تحليلية تطبيقية (*)

عرض: د. محمد كامل أحمد عبد الجواد

مدرس بقسم علوم المعلومات

كلية الآداب - جامعة بني سويف

(*) محمد كامل أحمد عبد الجواد. نظم استرجاع معلومات الويب المبنية على الإدراك: دراسة تحليلية تطبيقية؛ إشراف محمد جلال سيد غندور. سيد ربيع سيد -. كلية الآداب، قسم علوم المعلومات، ٢٠١٧. (أطروحة دكتوراه)

تمهيد

ساعدت شبكة الإنترنت العديد من الباحثين واخصائي المعلومات في الحصول علي معلومات هائلة في كافة المجالات. لكن الحجم الهائل والزيادة المستمرة في قواعد البيانات أصبحت تعيق المستفيد، ومن المعروف أن أدوات بحث الويب تتعامل حتى الوقت الحالي علي استرجاع المعلومات من منطلق الشكل؛ حيث أنها تعمل من خلال الأدوات المتوفرة لديها حالياً علي استرجاع الكلمة من خلال المطابقة الشكلية لكن لا يوجد لديها أدوات تتعامل مع استرجاع الكلمة بالمفهوم أو المعني؛ حيث أنها لا تتوفر لديها محللات صرفية ومعاجم لغوية. كما أنها لا تتعامل مع تقنيات محركات بحث الجيل الثالث من الويب، مع الوضع في الإعتبار أن الجيل الرابع من محركات بحث الويب علي وشك الظهور.

إن المعلومات النصية المتاحة علي شبكة الإنترنت يتم إتاحتها في أكثر من شكل سواء في شكل مكتوب أو في شكل مصور ولكي تعمل أدوات بحث الويب علي استرجاع هذه المعلومات من خلال المعني أو المفهوم فإن هذا يتطلب التعامل مع المحللات الصرفية والمعاجم اللغوية. لذلك فإننا في حاجة إلي النظر إلي بنية نظم الإدراك والتطبيقات التي تحتاجها والعمل علي وضع تصور عام لنظام الإدراك من خلال وضع تصور لشاشة عرض النظام، وكيف يكون عليه شكل البحث، وشكل النتيجة النهائية التي سوف تظهر أمام المستفيد. كما تعمل الدراسة علي وضع تصور نهائي لشكل المعالجة الفنية والاسترجاع بالمعني والمفهوم علي الويب

مشكلة الدراسة

تتمثل مشكلة الدراسة في زيادة تدفق ونمو حجم المعلومات الرقمية علي شبكة الإنترنت، وهذه المعلومات توجد في اشكال مختلفة كصفحات ومواقع وقواعد بيانات الويب، هذا بالإضافة الي انها قد توجد في اشكال مختلفة كالنص والصوت والصورة الثابتة والمتحركة، الي جانب تعدد موضوعاتها. بالإضافة الي ماسبق نجد أيضاً إختلاف نظم الإدراك في المعالجة الفنية والتشغيل لاسترجاع المعلومات

الرقمية؛ حيث اننا نجد أن هذه النظم تفتقر الي التعامل مع المحللات الصرفية والمعاجم اللغوية، كما أن نظم الإدراك تعمل من خلال مجموعة البرامج الفرعية ذاتية العمل وتختص بصفحات ومواقع الويب، كما أنها تعتمد علي العنصر البشري في اجراء عملية المعالجة والتشغيل لدليل البحث.

تساؤلات الدراسة

- كيف يمكن وضع تصور لنظام يعمل بكفاءة وفاعلية لتحقيق إدارة وتنظيم ومعالجة واسترجاع المعلومات النصية الرقمية المصورة المبنية علي الإدراك أو المعني في البيئة الرقمية العربية؟
- إلى أي مدى تتحقق في نظم إدراك المعلومات النصية المصورة في البيئة الرقمية العربية على الويب، الخصائص المميزة الواجب أن تتوافر فيها؟
- ما التقنيات اللازمة للتعامل مع معالجة وتكشيف النصوص الرقمية المصورة على الويب؟
- ما آليات ومعايير البحث اللازمة للتعامل مع نظم الإدراك للعمل على استرجاع وإدراك النصوص الرقمية المصورة على الويب؟
- ما الخصائص التفاعلية والديناميكية التي تتسم بها نظم إدراك المعلومات النصية الرقمية المصورة على الويب؟
- ما المواصفات التي ينبغي أن تتوافر في نظام إدارة وتنظيم ومعالجة واسترجاع المعلومات النصية المصورة من خلال المفهوم أو المعني في البيئة الرقمية العربية؟

أهمية الدراسة

- أن تزايد حجم مجموعات النصوص الرقمية المصورة المتاحة على الويب، وتزايد أهميتها، واستخدامها من قبل الباحثين في كافة الميادين البحثية

والتعليمية والإعلامية، يجد صدها بطبيعة الحال في البيئة الرقمية العربية، سواء من حيث العمل على تكوين مجموعات من هذه المصادر، أو من حيث اهتمام المستفيدين بالبحث عنها والوصول إليها للإفادة منها.

• محاولة التصدي بصورة شاملة لموضوع نظم استرجاع المعلومات الرقمية المبنية علي الإدراك، بوصفها وسيطاً للمعلومات، وبغية اكتشاف طبيعتها وخصائصها، وكيفية معالجتها، وتنظيمها، وكيفية إجراء عملية البحث والاسترجاع بالمفهوم أو المعني.

• تمثل نظم استرجاع المعلومات النصية الرقمية المبنية علي الإدراك شكلاً مختلفاً عن غيرها من نظم الاسترجاع الأخرى، لما تتسم به من خصوصية في عملية المعالجة والتنظيم وبحث واسترجاع المعلومات.

• اضافة نوعاً جديداً من نظم بحث الويب يكون قادراً على التعامل مع استرجاع وإدراك النصوص الرقمية المصورة، مما يساعد على سرعة بحث واسترجاع هذه المعلومات النصية المصورة.

• إن البحث عن مصادر المعلومات النصية الرقمية على الويب واسترجاعها من خلال المعني أو المفهوم، يقتضي توافر مجموعة من النظم المتخصصة التي تلاءم طبيعة وخصائص كل نوع من المصادر الرقمية سواء من حيث بنيتها، أو أساليب وتقنيات الوصول إليها ، أو طريقة عرضها، الأمر الذي يتطلب تحديد هذه الإمكانيات وبيان مدى توافرها فيما هو متاح من نظم البحث على الويب.

• سوف تركز الدراسة الضوء علي التحول والتبديل في المفاهيم وجوانب التنظير المختلفة التي تبني عليها نظم إدراك المعلومات النصية المصورة في البيئة الرقمية العربية.

- محاولة وضع تصور مقترح للمواصفات الواجب توافرها في أي نظام يعنى بإدراك واسترجاع المعلومات النصية الرقمية المصورة على الويب.

أهداف الدراسة

- وضع تصور لنظام يعمل بكفاءة وفاعلية لتحقيق إدارة وتنظيم ومعالجة واسترجاع المعلومات النصية الرقمية المصورة المبنية على الإدراك أو المعنى في البيئة الرقمية العربية.
- تحليل الخصائص المميزة لنظم إدراك المعلومات النصية المصورة في البيئة الرقمية العربية وتقييم آليات عملها.
- التعرف على التقنيات اللازمة للتعامل مع معالجة وتكثيف النصوص الرقمية المصورة على الويب.
- الوصول إلى آليات ومعايير البحث اللازمة للتعامل مع نظم الإدراك، وذلك حتى يمكن إدراك واسترجاع النصوص الرقمية المصورة على الويب.
- تحليل الخصائص التفاعلية والديناميكية التي تتسم بها نظم إدراك النصوص الرقمية المصورة في بيئة الويب.
- وضع تصور مقترح لمواصفات نظام إدراك واسترجاع المعلومات النصية الرقمية المصورة في البيئة الرقمية العربية.

منهج الدراسة

اعتمدت الدراسة على كل من:

- المنهج الوصفي التحليلي: وذلك لتتبع مجموعات الملفات النصية الرقمية المصورة على الويب، ورصد الآليات التي تعمل بها نظم إدراك النصوص الرقمية المصورة في معالجة ملفاتها الرقمية واسترجاعها، بالإضافة إلى استقراء الانتاج الفكري والدراسات المرتبطة بهذه الظاهرة.

- المنهج التجريبي: وذلك لملاحظة ومشاهدة واقع العمل بين نظم بحث العينة المختصة بإدراك النصوص الرقمية المصورة على الويب، وخصائص العمل بها، من خلال ادخال بعض المصطلحات البحثية ومقارنة النتائج بين نظم الإدراك المختلفة، وآليات عرضها على المستخدمين.

محتويات الدراسة

الفصل الأول بعنوان: معالجة النصوص في بيئة الويب. وقد تناول معالجة النصوص في البيئة الرقمية من حيث التعريف بها، ونشأتها، وتطورها التاريخي، وخصائصها، وأشكالها، وقضايا النص الرقمي، بالإضافة إلى العقبات والمشكلات الخاصة بمعالجة النصوص العربية بهدف توفير بعض الحلول للمشكلات المفتوحة التي يجب حلها من أجل تحويل اللغة العربية من لغة شحيحة المصادر إلى لغة مؤصلة ومدروسة من وجهة النظر الحاسوبية .

الفصل الثاني بعنوان: الإدراك لمعلومات الويب، وتناول الاتجاهات العامة للإدراك في بيئة الويب، بالإضافة إلى الأساليب المتبعة لإدراك المعلومات النصية سواء كانت الرقمية أو النصوص غير الرقمية، كما تناول المستويات الخاصة بإدراك المعلومات الرقمية سواء من خلال المستوي الشكلي للنصوص أو المستوي اللغوي، والمستوي التحليلي. وأخيراً تناول الفصل آليات الإدراك لاسترجاع المعلومات النصية والمراحل الرئيسية التي يمر بها نظام ال OCR. وأخيراً تطبيقات الإدراك الآلي للنصوص الرقمية في بيئة الويب.

الفصل الثالث بعنوان: نظم المعلومات المبنية على الإدراك. وتناول أهم التقنيات والبرمجيات الرئيسية المستخدمة في عملية إدراك الكتابات النصية العربية، والتي تمثل مجموعة من المقومات والدعائم التي تساعد نظام الاسترجاع بالإدراك للقيام بمهامه بشكل جيد، بالإضافة إلى البنية العامة لنظام الإدراك وكيفية الوصول للنصوص الرقمية المصورة، وآليات البحث المستخدمة في نظام

البحث، وايضاً التحديات والصعوبات التي تواجه عملية البحث والادراك، وأخيراً تقنيات وعلميات الإدراك وعلاقتها بانطولوجيا الادراك.

الفصل الرابع بعنوان: أدوات البحث بالإدراك في بيئة الويب. وتناول الفصل الخصائص المميزة لنظم إدراك النصوص الرقمية المصورة على الويب، وذلك من خلال تحليل وقياس كفاءة نظم إدراك النصوص الرقمية المصورة.

الفصل الخامس بعنوان: تصور مقترح لنظم بحث النصوص المبني على الادراك في البيئة الرقمية العربية. وقد اقتصر هذا الفصل على عرض المواصفات المقترحة لبناء نظم إدارة واسترجاع النصوص الرقمية المصورة في بيئة الويب، والخصائص المميزة لها، وذلك ابتداء من عمليات ادخل النص المصور إلى النظام مروراً بتنظيمه ثم محددات البحث والاستدعاء للنص المصور باللغة العربية والتي يمكن من خلالها تحقيق أفضل درجة لنتائج البحث أمام المستخدمين. وأخيراً واجهات التعامل للنصوص الرقمية المصورة في البيئة الرقمية العربية.

نتائج الدراسة:

وانتهت الدراسة بمجموعة من النتائج والتي كان من أهمها ما يلي:

- أظهرت الدراسة أن مجال إدراك النصوص الرقمية المصورة يعد مجال حديث نسبياً ولا يزال البحث في هذا المجال مستمراً، ويرجع ذلك لعدم كفاءة أساليب وآليات البحث الخاصة باللغة العربية، والتي تأخذ في عين الإعتبار كافة خصائص اللغة العربية. بالإضافة إلى أن محركات بحث الويب تتعامل مع الكم غير المحدود من المعلومات غير المعيارية البناء الرقمي، مما يؤدي إلى عدم وجود آليات لبحث واسترجاع النصوص الرقمية المصورة، فضلاً عن معياريها وقدرتها على التعامل مع مثل هذه النوع من الملفات الرقمية على الويب.
- أثبتت الدراسة أن إدارة محتوى النصوص الرقمية تنتمي أكثر إلى البيئة الرقمية الغربية في تنفيذ عمليات إدراك وتحليل المحتوى للنصوص المصورة، وقد

إنعكس ذلك بدوره على مدى التقصير في إجراء وتنفيذ عمليات المعالجة والتكشيف للنصوص الرقمية المصورة في البيئة الرقمية العربية.

- أظهر إختبار آليات بحث اللغة العربية مدى إفتقار جميع نظم إدراك النصوص المصورة بعينة الدراسة في إستخدام تقنيات التعامل مع اللغة العربية، بما يدل على غياب أدوات معالجتها؛ حيث أن عمليات التكشيف والتحليل لمحتوى النصوص الرقمية العربية تحتاج إلى بناء معاجم لغوية ومحللات صرفية لا تتمتع فقط بالقدرة على تحليل الكلمات لغوياً، وإنما القدرة على فهم وتحليل المحتوى الفكري للكتابات النصية العربية.

- أوضحت الدراسة أن عملية الإدراك Recognition هي العملية الأساسية التي يتم من خلالها نجاح عملية التكشيف والتحليل للكلمات النصية المصورة لكي ينتج من خلالها مجموعة من المعانى والمفاهيم المحددة التي تستخدم ككلمات ومصطلحات تعبر عن المحتوى الفكري للنص المصور.

- قدمت الدراسة في فصلها الأخير مجموعة من التصورات والمواصفات التي يجب أن تتحلّى بها نظم إدراك النصوص الرقمية المصورة على الويب ، حتى يمكنها فيما بعد العمل بكفاءة مع هذا الشكل من المصادر المصورة من حيث؛ المعالجة، والتنظيم، والبحث، والاسترجاع، وإدارتها، وإتاحتها للاستخدام أمام المستخدمين.

توصيات الدراسة

عمد الباحث إلى مجموعة من التوصيات منها ما يلي:

- أن تبني الشركات المنتجة والداعمة لملفات النصوص الرقمية المصورة معيار وصف لإنتاج ملفات النصوص المصورة بالشكل الذي يتيح لنظم إدراك النصوص الرقمية المصورة تكوين مجموعات من الصور النصية بشكل يسهل عملية البحث، فقد افتقدت دراستنا الحالية لنظم بحث وإدراك النصوص الرقمية المصورة إلى معايير ومقاييس علمية مقننة.

- لابد من تضافر الجهود العربية من أجل انشاء نظام قاموس حاسوبي عربي يتم من خلاله تصنيف وتنظيم قواعد الصرف والقواعد النحوية، وذلك حتى يمكن التغلب على المشكلات اللغوية الخاصة باللغة العربية ومنها على سبيل المثال مشكلة الترادف.
- يجب أن تسعى جميع نظم استرجاع المعلومات العربية إلى دعم خصائص اللغة العربية في عملية البحث وذلك ومن خلال توفير الخيارات المختلفة التي تدعم عملية البحث عن مشتقات الكلمة المدخلة في صيغة البحث، وعن الكلمات التي تنتج عن الكلمة المطلوبة، بعد أن تضاف إليها اللواحق واللواحق.
- يجب على الحكومات العربية وضع سياسة واضحة لإنتاج المحتوى المعرفي العربي رقمياً. وذلك من خلال وضع مجموعة من القوانين والتشريعات والقوانين ، وذلك لضبط تداول ملفات النصوص الرقمية العربية المصورة على الويب التي تتعامل مع مواد المعلومات وأدوات بحثها.
- يجب على أقسام علوم المعلومات والمكتبات الإهتمام بالتطورات السريعة التي تتم على مواد المعلومات النصية على شبكة الويب، ومن ثم العمل على وضع وتوفير الأجهزة والنظم والآليات اللازمة لبحث واسترجاع مثل هذه النوعية من مواد المعلومات على الويب في البيئة الرقمية العربية.