

Dirassat & Abhath
The Arabic Journal of Human
and Social Sciences



مجلة دراسات وأبحاث
المجلة العربية في العلوم الإنسانية
والاجتماعية

EISSN: 2253-0363
ISSN : 1112-9751

محركات البحث بين ضرورات التنظيم وتحديات الوصول إلى المحتوى العربي

**Search engines between the necessities of organization and the challenges of
access to the Arabic content**

كمال بطوش

kamel battouche

battouche_kamel@yahoo.com

جامعة عبد الحميد مهري قسنطينة

University of Abdel Hamid Mahri Constantine

مخبر بحث: طريق الجزائر نحو مجتمع المعلومات

Algeria's way towards the information society Laboratory

غنية زايدي

ghania zaidi

ist_jijel@yahoo.fr

جامعة جيجل

University of Jijel

مخبر بحث: طريق الجزائر نحو مجتمع المعلومات

Algeria's way towards the information society Laboratory

تاريخ الاستلام : 2018-09-17

تاريخ القبول : 2018-11-22

ملخص :

تربعت محركات البحث على عرش نظم بحث واسترجاع العنكبوتية بفضل عملها الآلي المتميز وتغطيتها الشاسعة للويب وحسن تنظيم واسترجاع موارد الشبكة. وبرغم التطور الذي عرفته محركات البحث حتى تتلاءم مع بيئة الويب المتطورة وتتمكن من تنظيمها وإتاحة الوصول إلى محتواها التراكمي بجميع لغاته، لا زالت العديد من التحديات تعرقل أداءها خاصة عند تعاملها مع المحتوى العربي ذا الطبيعة التركيبية الخاصة والبناء اللغوي المعقد. وتظهر هذه الورقة أهمية هذه الوسائل في تأمين أحسن وصول لمحتوى العنكبوتية وتنظيمه، وأهم تحديات التعامل مع المحتوى العربي.

الكلمات المفتاحية: محركات البحث؛ المحتوى العربي؛ الويب؛ تحديات الوصول؛ تنظيم المحتوى.

Abstract:

The search engines are ranked on the throne of search and retrieval systems of the web thanks to its excellent automation and extensive coverage of the Web and good organization and retrieval of network resources. Despite the evolution of search engines to adapt to the sophisticated web environment and to be able to organize and allow access to its cumulative content in all languages, there are several challenges that hinder their performance, especially when dealing with complex Arabic content. This paper shows the importance of these tools in securing access to the content and organization of the web, and the most important challenges dealing with the Arab content.

Keywords: Search engines; Arabic content; the web; access challenges; Organize content.

مقدمة:

لقد عدت شبكة الأنترنت أو بالأحرى بيئة الويب الرافد الأساسي للمعلومات، والبيئة الحاضنة للمحتوى الشبكي المتنامي باختلاف لغاته، أشكاله، طبيعته وخصائصه... وكان لزاما إيجاد وسائل وآليات دعم تواجد هذا المحتوى وإتاحة سبل تنظيمه والوصول إليه، وقد اعتلت "محركات البحث" قمة هذه الوسائل إن لم تكن أهمها، بفضل مرونة نشاطها وكفاءة أداؤها في مجال بحث وتقصي الويب. فوظفت هذه الأدوات كنظم أساسية لاستقطاب موارد الويب والعمل على تنظيمها وتبويبها في شكل صفحات ومواقع وتأمين استرجاعها والإفادة منها.

وقد وضعت الطبيعة الخاصة لهذه البيئة وما أضفته من الحرية والديمقراطية في مسائل نشر المعلومات بحثها واسترجاعها والمشاركة فيها، محركات البحث أمام تحديات كبيرة في مجال ضبط وتنظيم المعلومات وإتاحة الوصول إليها. وإن كان الحديث عن ضرورات التنظيم وتحديات الوصول للمحتوى الشبكي عموما يكتسي أهمية بالغة، فتخصيص الحديث عن المحتوى العربي لا يقل شأنًا عن ذلك، بل ربما جاز أن نوليه أكبر الإهتمام نتيجة جملة الخصائص التي طبعت المحتوى العربي وبنائه الخاص ومختلف الإشكاليات التي يطرحها في مجال التعامل معه والوصول إليه من قبل محركات البحث. وتحاول هذه الورقة تسليط الضوء على آليات تعامل محركات البحث مع المحتوى الشبكي عموما والمحتوى العربي بصفة خاصة في مجالات التنظيم وإتاحة الوصول إليه في ظل عديد المسائل المؤثرة. مجيبة على التساؤل الرئيسي التالي: ما دور محركات البحث في تنظيم المحتوى العربي وإتاحة الوصول إليه عبر الويب وما التحديات التي تصادفها؟

وسوف تحاول هذه الورقة البحثية تسليط الضوء على آليات تعامل محركات البحث مع المحتوى الشبكي عموما والمحتوى العربي بصفة خاصة في مجالات التنظيم والوصول إليه في ظل عديد المسائل المؤثرة.

فرضيات الدراسة: تحاول من خلال هذه الدراسة اختبار مجموعة الفرضيات التالية:

✓ **فرضية 01:** تزداد أهمية محركات البحث في تنظيم المحتوى الشبكي على اختلاف لغاته في ظل تنامي بيئة الويب حجما، شكلا، مضمونا وتفاعلا.

✓ **فرضية 02:** تصادف محركات البحث عديد التحديات في مجال تنظيم وإتاحة الوصول للمحتوى الشبكي ترتبط بجميع مراحل عمل المحرك.

✓ **فرضية 03:** يطرح المحتوى العربي تحديات في مجال تنظيمه وإتاحة استرجاعه عبر المحركات نتيجة طبيعته التركيبية الخاصة وبناءه المعقد ما يستدعي آليات فهم هذا المحتوى تركيبيا، بناءا ودلالة.

أهمية الدراسة :

تنبع أهمية الدراسة من أهمية المعلومات في البيئة الإلكترونية في حد ذاتها وأهمية تيسير سبل الوصول إلى المحتويات الشبكية خاصة باللغة العربية، لاسيما وأنه لا جدوى من وجود هذا المحتوى إذا لم تتوفر قنوات وآليات تتيح الوصول إليه وتنظيمه قصد تحقيق أفضل إفادة منه.

كذلك أهمية دعم تواجد اللغة العربية على الويب حتى لا يتلاشى المحتوى العربي أمام باقي المحتويات، وأهمية التعرف على خصائص اللغة العربية وبناء محركات بحث متطورة تراعي بناءها وتركيبها ودلالاتها حتى لا يبقى محتوانا العربي مهمشا.

أهداف الدراسة: يمكن إجمال أهم الأهداف التي تسعى إليها الدراسة في:

- التعرف على بيئة الويب باعتبارها البيئة الحاضنة للمحتوى الشبكي ولنشاط محركات البحث
- التعرف على محركات البحث كأهم نظم بحث واسترجاع العنكبوتية وأوجه تمييزها عن غيرها.
- استطلاع واقع المحتوى العربي على صفحات الويب وأهم خصائصه البنائية والتركيبية
- تحديد أهم صعوبات التعامل مع المحتوى العربي من قبل محركات البحث تنظيما، تكييفيا وترتيبيا.

منهج الدراسة: إغتمدنا في الدراسة على المنهج الوصفي لأنه الأنسب لتحقيق أهدافها، فهو الذي يقوم على أساس تحديد خصائص الظاهرة ووصف طبيعتها ونوعية العلاقة بين متغيراتها وأسبابها واتجاهاتها.

1- بيئة الويب: أي فضاء للمعلومات؟

جسدت الويب منذ وضع أول خادم لها في العالم من قبل البريطاني تيم برنرزي في بداية 1990¹، الإنطلاقة الحقيقية للأنترنت

- ✓ انعدام ضمانات دوام المواد على الويب وعدم اختفائها يشكل انشغالا كبيرا للأفراد المهتمة.
- ✓ زيادة هائلة ومستمرة لأحجام كبيرة للمعلومات على خدمات الويب.
- ✓ ابتكارات تكنولوجية دون انقطاع تشهدها الشبكة العنكبوتية العالمية
- وتعد هذه الخصائص مؤشرات هامة على تعقد مهام وسائل ضبط وإتاحة الوصول إلى معلومات الشبكة.
- 2- محركات البحث كنظم استرجاع العنكبوتية.

تربعت محركات البحث على عرش نظم استرجاع الويب وقد غدت أكثر تطبيقات الأنترنت شيوعا وأبرز أدوات البحث استخداما مقارنة بالأدلة والبوابات.. ويتولى نظام استرجاع المعلومات مهام تمثيل وحدات المعلومات وتخزينها وتنظيمها وتسهيل الوصول إليها. و سواء كانت إلكترونية أو تقليدية فهي بمثابة وسيط بين المستفيد والمعلومات النهائية⁰⁵، وتظهر أهميتها في بيئة الويب المتطورة نتيجة صعوبات البحث عبرها والتي تشمل:

- ✓ الصعوبات المتعلقة بالبيانات وطبيعة الشبكة العنكبوتية .
- ✓ ما يتعلق بالمستفيد وتفاعله مع النظام .
- 1.2- تعريف محركات البحث:

هي أدوات تعمل بشكل آلي، ظهرت لتنظيم معلومات الويب في مختلف أشكالها وضمان الوصول إليها. ويبرز عمل محركات البحث من خلال كونها أدوات تسيح في فضاء العنكبوتية الواسع لتلتقط كل مصادر المعلومات باختلاف مجالاتها وأشكالها، وتعتمد على ثلاث برامج أساسية: الزاحف المكشف وآليات البحث بما يوفر القدرة على الإضافة والبحث⁰⁶ وهي أيضا برمجيات ويب تسمح بالعثور على المصادر المختلفة بالاستعانة بالكلمات المرتبطة بها وفق ثلاث مهام أساسية:⁰⁷

التصفح: يتم بزيارة صفحات ومواقع الويب أليا عن طريق الروبوت.

التكشيف الآلي: القيام بتحليل الصفحات لاستخراج الكلمات الدالة.

خاصة وأنها لم تكتفي بالهدف الأولي الذي حدد لها وهو إتاحة تبادل المعلومات بشكل فوري، بل أعطت بعدا جديدا للمشاركة في المعلومات والمعارف وتداولها بشكل حروديموقراطي فضلا عن إتاحتها لعدد التقنيات في مجال عرض المعلومات ودمجها باعتمادها على تقنية النص الفائق ولغة (html:hyper text markup langage) لتعتمد فيما بعد على نظم إدارة المحتوى ولغات ترميز متطورة مكنت من تطوير الويب إلى شبكة عالمية أتاحت مرونة وديناميكية في عرض المعلومات فضلا عن عديد التطورات تجسدت عبر عدة أجيال مرور الويب 1.0 فالويب 2.0 أو الويب الإجتماعي ثم الويب 3.0 أو الويب الدلالي، وهناك حديث عن ويب رابع web4.0 تشهد إرهاساته حاليا². وقد امتازت الويب كمنصة عمل على الأنترنت بمجموعة عناصر أهمها:³

- البساطة في العمارة حيث الاعتماد على النص الفائق وعلى محدد URL
- ✓ التشبيك: حيث النمو الشبكي المرن في المحتوى و الخدمات.
- ✓ القابلية في التوسع: بداية بويب الوثائق إلى غاية الويب الدلالي.
- ✓ العالمية: باعتماد نسق وأنظمة متباينة وتعددية في لغات محتوياتها.
- ✓ التكلفة: حيث انخفاض تكلفة التشغيل من متصفحات مجانية وخدمات واتصال بالانترنت.
- ✓ التنوع الفكري: باشتغالها على أغراض متعددة من تثقيف وترفيه وتعليم.
- ولقد أوجدت الويب بيئة مختلفة للبحث عن تلك المتاحة عبر قواعد البيانات بفضل تميزها ب:⁴
- ✓ التنوع وعدم التجانس من حيث طبيعة المحتوى أو طرق عرضه أو مصادره أو حتى الغرض من نشره .
- ✓ نقص الرقابة والتنظيم فلا وجود لهيئة مركزية للتنظيم، ولا جهاز رقابة.
- ✓ تعدد التشريعات المطبقة على وسائل الإعلام العابرة للحدود.

تظهر أهمية المحركات من خلال جملة المهام التي تؤديها ومنها: ¹⁰	المساءلة: عن طريق مطابقة كلمات تساؤل المستفيد بالكلمات المكشوفة المخزنة داخل المحرك وتعرض النتائج وفق درجة ملائمتها.
✓ جمع وإضافة المعلومات في أشكالها المتعددة وباختلاف موضوعاتها.	2.2 مكونات محركات البحث وطريقة عملها: تحتاج المحركات لأداء دور الوساطة إلى ⁰⁸ :
✓ القيام بعمليات التنظيم والمعالجة للمعلومات على اختلاف مستوياتها، وإتاحة الوصول إليها.	✓ نظام للتجميع بغرض ملء قاعدة البيانات بالمعلومات.
✓ توفير الموارد البرمجية التي تساعد على تحقيق أفضل النتائج .	✓ نظام للتكشيف لتنظيم محتويات قاعدة البيانات.
✓ العمل على التطوير والتحديث الدائم للمعلومات المخزنة و لاستراتيجيات ولغات البحث.	✓ خوارزمية للبحث في قاعدة البيانات.
✓ تقديم المعلومات المسترجعة والنتائج بما يلائم حاجة المستفيد.	✓ خوارزمية للترتيب وتنظيم قائمة النتائج.
وتنفرد محركات البحث بأعلى نسب استخدام كما تستحوذ على أكبر عدد من الإستفسارات بلغت نحو 150 مليون استفسار في اليوم الواحد، وتكشف نحو 16% من محتوى العنكبوتية ¹¹ . وقد سعت محركات البحث إلى تطوير أدائها بما يتلاءم وطبيعة البيئة الشبكية المتطورة، وظهرت محركات بمزايا جديدة نذكر منها: ¹²	ويتسنى لها أداء مهامها من خلال بنائها الهيكلي الذي يضم البرامج: ⁰⁹
الشمول: أي القدرة على البحث في مختلف الموارد الشبكية في وقت واحد وبشكل متزامن بدلا من استخدام محرك بحث لكل فئة على حدى.	✓ الروبوت robot: يسمى أيضا الزاحف أو العنكبوت crawler- spider ويقوم باستكشاف الويب لإيجاد الصفحات الجديدة عن طريق تتبع الروابط و التنقل بين المواقع. ويشكل قاعدة بيانات ضخمة بالمواقع التي قام بزيارتها، ويحدثها دوريا.
السرعة: أي تحصيل أفضل سرعات الإستجابة وعرض النتائج.	✓ الكشاف index أو المفهرس: يقوم الروبوت بإرسال البيانات والمعلومات المجمعة عن المواقع إلى أداة التكشيف من أجل تكوين كشاف بالمصطلحات التي جمعها الروبوت ويقوم باختزائها في قاعدة بيانات المحرك مع ربط كل مصطلح بعنوان الصفحة المتعلقة به حتى يسهل الوصول إليها.
العمق: أي إمكانية فهم المحرك لمعاني ودلالات الكلمات .	✓ أداة البحث searcher: تسمح بصياغة الإستفسار واستعراض النتائج في شكل قائمة. فبعد تلقي الإستفسار يتم البحث في كشاف المحرك ، للحصول على المواقع المتطابقة مع الإستفسار وتعرض بعدها النتائج وفق ترتيب معين.
الشخصنة: أي إتاحة عملية البحث وفق احتياجات وتفضيلات المستعلم.	3.2 أهمية محركات البحث كنظم استرجاع المعلومات:
3- محركات البحث وتنظيم المعلومات:	
لقد أثبتت محركات البحث جدواها في مجال تنظيم المعلومات الشبكية من خلال إسهامها في فحص الويب وتنظيم موارده عن طريق تقصي المواقع، تكشيف النصوص وبناء كشافات ضمن قواعدها والتعرف على مختلف المواد واستعراض النتائج بفضل آليات وبرامج معينة تتضمنها. وتطلع محركات البحث لتنظيم معلومات الويب بمهام: ¹³	

برغم تضاعف المحتوى العربي على شبكة الأنترنت ظلت مساهمته في إثراء المحتوى الشبكي متواضعة، ولا ترقى أبداً إلى مكانة اللغة العربية ولا إلى عدد المتحدثين باللغة العربية على الأنترنت، ولا حتى إلى أعداد مستخدمي الأنترنت العرب التي شهدت نمواً هائلاً في السنوات الأخيرة، وتشير الكثير من الإحصائيات إلى تدني المحتوى العربي كما وكيفاً وتصفيحاً، وقدرته إحدى الدراسات بـ 3% من إجمالي المحتوى العالمي¹⁵.

ولا تقل أهمية المحتوى العربي الرقمي عن أهمية المحتوى الشبكي عموماً في كونه جوهر العنكبوتية ومضمونها الفكري، المعرفي، الترفيهي، التواصلية.. وهو أساس وجود الأطراف الأخرى للشبكة، فضلاً عن مساهمته في خدمة الجوانب العلمية والثقافية والتكنولوجية لمنتجي وناشري ومستعملي هذا المحتوى، وتنمية مواردهم البشرية وتعميق البحث والتطوير. وقد طرحت حقيقة الوضع الذي يشهده المحتوى العربي الرقمي عديد الإشكاليات تنوعت بتنوع وتعدد مراحل تهيئته واختلاف الفاعلين والمساهمين فيه والمستفيدين منه وكذا تطور تقنيات بثه ومعالجته والوصول إليه... إضافة إلى ارتباطه باللغة العربية وهي لوحدها قضية متعددة الوجوه والإشكاليات نطقاً، كتابةً، إحصائياً... بما يرتبط بطبيعة اللغة العربية نفسها وبخصائصها المعلوماتية¹⁶. وهي قضايا هامة تؤثر على عمل محركات البحث.

فإضافة إلى أن المحتوى الرقمي العربي لم يكن له الحظ الأوفر في الحضور على صفحات الويب، فقد افتقرت الويب ومحركات البحث خلال مدد طويلة إلى التعامل مع مواد المعلومات المكتوبة باللغة العربية نظراً لنقص أو انعدام برمجيات وآليات البحث الداعمة للتعامل مع اللغة العربية، خاصة وأنها مواد تتطلب تعاملاً خاصاً وتتجلى الصعوبات التي واجهها المحتوى العربي في عمليات البحث والإسترجاع على صعيدين إثنين¹⁷:

✓ صعوبة البحث بالمحركات الأجنبية: لأنها مصممة خصيصاً للغة الأم.

✓ قلة فعالية المحركات العربية وهي أصلاً ذات واجهات عربية فقط.

وقد غدى هذه الصعوبات الطابع البنائي والتركيبية للغة العربية التي تعتبر أقدم اللغات عالمياً وسادس لغة معترف بها على مستوى الأمم المتحدة فهي أعقد اللغات السامية وأغناها صوتاً وصرفاً ومعجماً، وبخلاف اللغات اللاتينية، للغة العربية خصائص دلالية وتركيبية معقدة تؤثر على دقة وفعالية الإسترجاع¹⁸ مما يستوجب نظم بحث وإسترجاع

✓ الإضافة والإقتناء: بالإعتماد على الزواحف الآلية التي تجوب الصفحات والمواقع تقوم بزيارات دورية لتحديث قاعدة البيانات.

✓ المعالجة والتنظيم: ببناء التسجيلات آلياً عن طريق تجميع كلمات معبرة عن الموضوع تضمها صفحة الويب أو تخزين كامل النص مرفق بمكان تواجدته على الشبكة حتى يتم إسترجاعه.

✓ البحث والإسترجاع: تتيح المحركات البحث الحر في كامل نص الموقع أو أماكن محددة كالعنوان أو الوسوم.. ويكون ترتيب النتائج بحسب درجة الصلة بين الموقع والموضوع، ويأخذ بعين الإعتبار تكرار الكلمات المفتاحية للتساؤل أو أماكن تواجدتها داخل الموقع.

وبرغم العمل الآلي لمحركات البحث فهي تعجز عن التغطية الشاملة والوصول لكافة جوانب الشبكة وتؤكد الدراسات أن نسبة الإسترجاع على شبكة الويب لا تتعدى 15% أي أن احتمال 85% من المصادر لا يمكن إيجادها على الرغم من التطور الذي تشهده التقنية ما يشير إلى الفوضى التي تتسم بها عملية النشر وتنظيم الويب. ومن الصعوبات التي تواجه محركات البحث في تنظيم محتوى الشبكة: ¹⁴

✓ ضخامة مواد الويب واستمرارية نموها خاصة بظهور الويب الإجتماعي وأدوات النشر الحر.

✓ تنوع مصادر المحتوى عبر الويب، واختلاف طبيعتها وشكليتها وحتى طرق بلوغها.

✓ إختلاف لغات محتوى الويب، وصعوبة تعامل محركات البحث مع الحرف غير اللاتيني.

✓ التغير الدائم والديناميكية التي تشهدها مواقع وصفحات الويب.

ويبقى تنظيم الويب برغم عديد الإشكالات التي يطرحها، غاية ومسعى أساسي لمختصي المعلومات ومصممي أدوات، لذلك فجهود تنظيم البيئة الشبكية تظل قائمة وتطوير محركات بحث قادرة على الإسهام في المجال ضرورة ملحة.

4- المحتوى العربي على الويب بين إشكالات التواجد وطبيعة البناء والتركيب:

حدث من فاعلية أداء الأدوات التي أتاحتها لضبط وتنظيم المحتوى عبر جنباتها، فطبيعة الشبكة العنكبوتية ذاتها أثرت بشكل كبير في قضايا تنظيم واسترجاع المحتوى عموماً ومنه المحتوى العربي. إذ أنها تتميز بمحدودية معالجتها لدلالة المحتوى وعدم وجود تنسيق وتمثيل جيد للبيانات، فروابط الصفحات والمواقع لا تحمل أية دلالة سوى أن هذا الموقع مرتبط بذلك، وتتمثل أهم خصائص العنكبوتية كبيئة لتنظيم واسترجاع المحتوى في:²¹

✓ افتقارها للبنية السليمة لتمثيل المحتوى.

✓ الغموض المعلوماتي الناتج عن ضعف منهجيات الربط بين المحتوى.

✓ قابلية الاستخدام وبالتالي مسائل التعامل مع العدد الهائل من المستفيدين والحجم الهائل من المحتوى.

✓ افتقار عنصر المصداقية والثقة فيما تحمله العنكبوتية من معلومات.

✓ افتقار آليات وخوارزميات العنكبوتية من الفهم المنطقي للمحتوى

كما تطرح محركات البحث ذاتها تحدياً آخر يكمن في طبيعة أدوارها، حيث أنها صممت للإجابة على تساؤل واحد: ما هي الوثائق التي تشمل على الكلمات أو الجمل التالية؟ دون القدرة على التنبؤ بالوثائق ذات الصلة بموضوع الإستفسار وتحديد غير ذات الصلة. ويجمل البعض الآخر كافة التحديات في نطاق عاملين اثنين:²²

✓ المعالجة اللغوية داخل المحركات تركز على وجود أو غياب الكلمات المفتاحية دون تحليل للمحتوى أو فهمه ما يؤثر على دقة الاسترجاع.

✓ محدودية الآليات والخوارزميات في صياغة تعابير الإستفسار بشكل لا يساعد المحرك في التحديد الدقيق لما يريده المستفيد .

تحدي آخر تواجهه محركات البحث هو اختلاف طرق تشفير الوثائق وعدم دقة تمثيل المحتوى باستعمال لغات التشفير وواصفات الميتاداتا.. وتزخر الويب بعديد الوثائق المختلفة البناء والتركيب الشكلي باعتماد نوعيات مختلفة من برامج تحرير المواقع ونظم إدارة المحتوى، والتي تتيح جودة عالية في العرض مقابل افتقادها إلى فهم دلالات ما تعرضه، إضافة إلى اعتماد لغة النص الفائق والتي تقف عتبة عسيرة في

مجهزة بالبرمجيات والحلول المناسبة لاسترجاع المعلومات باللغة العربية. وبشكل عام تمتاز النصوص العربية على العنكبوتية العالمية ب:¹⁹

✓ أن النظام الإملائي للغة العربية يقوم على ثمانية وعشرين حرفاً، وقد يأخذ الحرف الواحد أشكالاً متنوعة.

✓ اختلاف شكل الحرف حسب موقعه من المقطع.

✓ تقرأ وتكتب نصوص اللغة العربية من اليمين إلى اليسار.

✓ هناك حروف تكتب ولا تنطق مثل همزة الوصل، وحروف تنطق ولا تكتب مثل ألف المد.

✓ الكلمات العربية مركبة تتضمن: التذكير والتأنيث، الزمن، العدد..

✓ تكتب الحروف العربية بأنماط مختلفة بحسب الخطوط العربية:النسخ، الرقعة، الثلث، الكوفي، الديواني...

✓ إختلاف الإمتداد الأفقي للحروف العربية ووجود تشابه كبير بين أشكالها

✓ وجود علامات الضبط بالشكل مما يؤثر على فهم الكلمة

✓ اتصال الكلمة بحروف السوابق واللواحق.

✓ يتفاوت إملاء الكلمات وتشكيلها بسبب الأخطاء البشرية واختلاف قواعد إملاء الكلمات المعربة والأجنبية.

✓ إختلاف في شكل الأرقام بين المشرق والمغرب العربي وأروبا .

ومن خصائص اللغة العربية التي تؤثر في جوانب المعالجة الآلية لها كما أجملها نبيل علي نذكر²⁰:

التوسط اللغوي، حدة الخاصية الصرفية، المرونة النحوية، الانتظام الصوتي، ظاهرة الإعراب، الحساسية السياقية، تعدد طرق الكتابة وغياب عناصر التشكيل، ثراء المعجم واعتماده على الجذور، شدة التماسك بين عناصر المنظومة اللغوية... ومثل هذه الخصائص وأخرى وإن كانت تدل على ثراء اللغة العربية، فهي من جهة أخرى تطرح تحديات كبيرة في مسائل بحثها وتنظيمها واسترجاعها وتسهيل الوصول إليها.

5- تحديات محركات البحث في تنظيم المحتوى العربي.

تعد قضايا تنظيم المحتوى عبر بيئة الويب من الأمور التي بالكاد يستحيل اكتمالها، ذلك أن هذه البيئة حملت بين طياتها عوامل

فالنسبة لزواحف محركات البحث، فهي تقف عاجزة أمام الأحجام الهائلة للمحتوى الشبكي، نتيجة محدودية الطاقة الإستيعابية للزواحف في مجال تجميع المحتوى. وقد أثبتت الدراسات أنه يستحيل على أي محرك بحث أن يقوم بتكشيف كلي للويب المرئي²⁶، إضافة إلى الإشكال المتعلق بمستوى التجميع الذي يجب التوقف عنده فضلا عن ارتفاع معدلات التغيير والتحديث للمواقع خلال فترات قصيرة فضلا عن صعوبة الوصول إلى الويب الخفي.

أما بالنسبة لتكشيف المحتوى فأول ما يواجهه محرك البحث هو اختلاف تنسيقات الكتابة ما بين صيغ pdf وhtml .. ما يستدعي تطبيع النصوص وفقا لشكل موحد وإعادة هيكلتها بصورة منطقية تمكن من البحث فيها، ويمكن إجمال التحديات التي تواجه تكشيف المحتوى العربي في محركات البحث في²⁷:

✓ ما تقوم به محركات البحث من عمليات تفتيت للمحتوى الكامل إلى كلمات مستقلة وهو لا يتناسب مع طبيعة بعض محتويات الوثائق .

✓ ما تقوم به المحركات من استبعاد لبعض الكلمات ذات دلالة ضئيلة ولكنها قد تلعب دورا محوريا.

✓ منهجية التكشيف داخل المحركات التي تقضي باستخراج الجذور الصرفية للمصطلحات الواردة في المحتوى دون أن تأخذ في اعتبارها أن اللغة العربية تتسم جذورها الصرفية بالتعقيد وعدم المرونة.

✓ مفاضلة المحركات بين صغر الكشاف وبين القدرة على الإجابة على الاستفسارات المعقدة.

✓ بناء الكشافات الفرعية ضمن الكشاف المقلوب مما يسمح بوجود تداخل وتكرار في النتائج .

كما تواجه محركات البحث تحديات في مجال ترتيب النتائج ورغم تطور خوارزميات الترتيب الطبقي داخلها، مثل تحدي التنبؤ بالمحتوى ذا الصلة بالموضوع وبرغم اعتماد طرق إحصائية لتحديد درجة صلة الوثائق بالإستفسار، مثل حساب تردد المصطلح، أو موقع المصطلح، تحليل الروابط... فلكل منها أوجه قصور خاصة في تعاملها مع اللفظ العربي ذا الثراء اللغوي والدلالات الغنية والتركيبات المعقدة. وهنا يجب النظر إلى الحلول التي توفرها بيئة الويب الدلالي وضرورة وجود أنطولوجيات عربية تستند عليها محركات البحث حتى تتمكن من فهم

تحليل مضمون الوثائق من قبل الزواحف، لأنها لا توفر توصيف للمحتوى بل هي تمثيل أصم للبيانات²³.

وتعد الطبيعة التركيبية والبنائية المعقدة للمحتوى العربي تحديا آخر أمام محركات البحث. فاللغة العربية تمتاز بكثرة التصرف والإشتقاق إذ تحمل حوالي خمسة مليون كلمة مشتقة من حوالي 11350 جذر مقارنة باللغة الإنجليزية التي تحتوي على 1.3 مليون كلمة، ويعود ذلك إلى وجود مئات الإشتقاقات من جذر واحد، وتعدد معاني الكلمة الواحدة واستخدام كلمات مختلفة لنفس المصطلح²⁴ ما يعد تحديا لتنظيم المحتوى من قبل محركات البحث .

6- تحديات محركات البحث في تكشيف المحتوى العربي وترتيبه:

تعد جودة تصميم محرك البحث وقدرة خوارزمياته على التعامل مع لغات المواد وأشكالها المحك الحقيقي لنجاح نظم وأساليب معالجة المعلومات وتكشيفها وتيسير استرجاعها. ولأن اللغة تعتمد على الأفكار والمفاهيم وهي ليست بالشيء الثابت فهي تعد تحديا بالنسبة لمحركات البحث في تكشيف نصوصها وضبط الوصول إليها، وتطرح اللغة العربية إشكالاتها في مجال التكشيف والضبط والوصول والإسترجاع نتيجة مرونتها واتساع مفرداتها وتعدد معانيها.. إضافة إلى مشكلات عدة تطرحها أمام محركات البحث أهمها²⁵:

✓ اشتغال الكلمات العربية على بعض السوابق(ال، ب ..) تؤدي إلى أن ترتب هذه الكلمات وفقا لها في الكشاف.

✓ التركيب الصرفي للكلمات في اللغة العربية.

✓ أحرف العلة في اللغة العربية.

✓ رد الكلمة لجذورها أو مادتها اللغوية للبحث عنها.

✓ مشكلات اختلاف البنية الصرفية للكلمة .

✓ مشكلة الشدة والحرف المزدوج.

وبرغم الإستعانة بما استجد في مجال تقنيات الويب الدلالي واعتماد أساليب تحسين إظهار المواقع، تبقى هناك عديد التحديات التي تواجه محركات البحث لتكشيف المحتوى العربي. وأهم التحديات تلك المرتبطة بعمل المحرك، سواء مع آليات الزحف، التكشيف أو الترتيب، والتي تعتبر تحديات داخلية.

✓ هناك حاجة ملحة لتطوير محركات بحث عربية تستند على تقنيات الويب الدلالي، وبناء أنطولوجيات عربية حتى يكون هناك فهم أعمق لدلالات الألفاظ والعلاقات فيما بينها وتقديم نتائج أكثر دقة.

تسعى محركات البحث إلى تنظيم المحتوى الشبكي وإتاحة أحسن وصول إليه وإن صادفت هذه المحركات عديد العراقيل والتحديات التي تزداد حدة عند التعامل مع المحتوى العربي فهذا لا يعني أن اللغة العربية قاصرة بل ينبغي تكاثف الجهود في مجال بناء محتوى رقمي عربي بما يسمح لمحركات البحث من الوصول إليه واستغلاله حتى لا يبقى مهمشا، فضلا عن ضرورة مراعاة تصميم مثل هذه الأدوات بما يتوافق مع طبيعة اللغة العربية والمشكلات التي يطرحها اللفظ العربي في المعنى أو المبنى.

الهوامش

¹ - delengaigne, xavier. Organiser sa veille sur internet : au de la 07de google. paris :eyrolles ;2012. P.

² - Accart Jean.philippe ;Rivier,Alescis. Mémento de l' information numérique.Paris:Editions du cercle de la librarié, 2012, p 51

³ - نشرتي، مؤمن سيد. الشبكة العنكبوتية الدلالية: هوية تبحث عن الوجود دراسة تأصيلية تحليلية. cybrarians journal. ع. 27 ديسمبر 2011 [على الخط] زيارة يوم 2015/11/15 متاحة على الموقع <http://journal-cybrarians.info/index.php?option=com-content&view=article&id=603.2011-12-01-22-24-43>

⁴ - lardy, jean-pierre.la recherche d'information sur internet : 2001.p.47 outils et méthodes.7ed. paris : adbs;

⁵ - إبراهيم، سيد ربيع. الويب: نظام استرجاع المعلومات الكوني. الرياض: مكتبة الملك فهد الوطنية. 2009. ص. 44.

⁶ - إبراهيم، سيد ربيع سيد. محركات بحث الوسائط المتعددة، المفهوم: الأداء الأنواع- Cybrarians Journal. ع 07 (2005) [على الخط] زيارة للموقع يوم 2015/09/05 متاحة على الرابط: <http://journal-cybrarians.info/index.php?option=com->

⁷ - denos, nathalie ; simili,karine. Organiser la recherche . Site visité [en ligne]d'information a l'ère numérique. Octobre 2011 - fiche / 11fr/ressources20 le.13/11/2016. Disponible sur: c2i.univ-littoral.d4-v1-0-20111020.pdf.

⁸ - الهوش، أبو بكر محمود. المرجع السابق. ص. 16.

⁹ - أحمد، أحمد فرج. دراسات في تحليل وتصميم مصادر المعلومات الرقمية. الرياض: مكتبة الملك فهد الوطنية. 2009. ص. 224- 228.

دلالات المعاني والعبارات داخل المحتوى العربي مثلما أثبتته عديد التجارب في المحتويات غير العربية.

الخلاصة:

لقد أوجدت البيئة الشبكية بمختلف متغيراتها عدة تحديات في مجال تنظيم والوصول إلى المحتوى. ويطرح المحتوى العربي زيادة على ذلك تحديات ترتبط بالطبيعة التركيبية المعقدة للغة العربية والتي تستدعي تعاملًا خاصًا من قبل محركات البحث. ويمكن إجمال نتائج الدراسة في:

✓ تطرح ضخامة الويب وتنوع موارده وديناميكيته..عديد التحديات في مجال تنظيم موارده وإتاحة الوصول إليها من قبل محركات البحث.

✓ تزداد أهمية محركات البحث في تنظيم موارد الويب المتزايدة والمتنوعة في ظل تطور أجيال الويب ما استدعى محركات بحث بخصائص جديدة مواكبة لمستجدات الساحة.

✓ يتزايد المحتوى العربي عبر صفحات العنكبوتية غير أنه يواجه صعابًا وتحديات لغوية، حاسوبية، معرفية وإدارية، وتطفو على السطح قضايا تعامل المحركات مع هذا المحتوى، معالجة، تنظيمًا وإسترجاعًا.

✓ كفاءة تعامل محركات البحث مع المحتوى العربي تتطلب فهم آليات بناء اللغة العربية ودلالات التراكيب اللغوية والنحوية والصيغ الإشتقاقية..

✓ نجاح محركات البحث في الوصول إلى المحتوى العربي وجودة تقديمه، رهن قدراتها على إدارة هذا المحتوى بتوفير المعالجات الفنية التي ينبغي أن تجرى على النصوص العربية وتوفر التحليلات المطلوبة لدعم المحرك.

✓ تتنوع أهم تحديات محركات البحث في تنظيم المحتوى العربي وتسهيل استرجاعه ما بين تحديات مرتبطة بزواحف المحرك، ببرمجيات التكشيف وبأساليب الترتيب.. وتجاوزها يستلزم مراعاة جميع مراحل عمل المحرك بما يتوافق مع خصائص اللغة العربية وبناءها ودلالاتها.

²⁴ - دبور، عبد الرحمن غالب. نظم استرجاع المعلومات العربية واتجاهات البحوث المعاصرة. مجلة إعلم، ع.14، فبراير 2015 [على الخط] زيارة يوم 2016/01/08 متاحة على موقع <http://arab-afli.org>.

²⁵ - النشترتي، مؤمن سيد. التحديات التي تواجه خوارزميات محركات البحث في استرجاع المحتوى العربي دراسة مسحية تحليلية. المرجع السابق.

²⁶ - foenix-rio, beatrice. Recherche éveillée sur internet : mode d'emploi.paris :lavoisier ;2011.p.19

²⁷ - النشترتي، مؤمن سيد. التحديات التي تواجه خوارزميات محركات البحث في استرجاع المحتوى العربي دراسة مسحية تحليلية. المرجع السابق.

قائمة المراجع :

1- إبراهيم، سيد ربيع سيد. الويب:نظام استرجاع المعلومات الكوني.الرياض:مكتبة الملك فهد الوطنية،2009.

2- إبراهيم، سيد ربيع سيد. محركات بحث الوسائط المتعددة، المفهوم: الأداء الأنواع- Cybrarians Journal - ع 07 (2005) [على الخط] زيارة للموقع يوم 2015/09/05 متاحة على الرابط: <http://journal.cybrarians.info/index.php?option=com->

3- إبراهيم، سيد ربيع سيد.نظم استرجاع قواعد بيانات الويب غير المرئية. الرياض: مكتبة الملك فهد الوطنية، 2010.

4- أحمد، أحمد فرج. دراسات في تحليل وتصميم مصادر المعلومات الرقمية. الرياض: مكتبة الملك فهد الوطنية، 2009.

5- الإسكوا. تعزيز المحتوى الرقمي العربي في المنطقة العربية.[على الخط]. زيارة للموقع يوم 2017/08/08 متاحة على: <http://css.escwa.org.IB/ictd/1791/04a.pdf>

6- الدكتور، أيمن شعبان. المحتوى العربي على العنكبوتية العالمية: منظور معلوماتي- Cybrarians Journal - ع 23 سبتمبر 2010 [على الخط]. زيارة يوم 2015/12/04 متاحة على الموقع: http://journal.cybrarians.info/index.php?option=com_content

7- الصويغ، علي بن سليمان. استرجاع المعلومات في اللغة العربية.ط.2. الرياض:مطبوعات مكتبة الملك فهد الوطنية، 2008 .

8- المعتم، نبيل بن عبد الرحمان. البحث باللغة العربية على محرك البحث جوجل google مجلة مكتبة الملك فهد الوطنية.مج.17، ع.2، نوفمبر 2011.[على الخط] زيارة يوم 2016/04/12 متاحة على الموقع www.kfnl.org.sa/Ar/mediacenter/EMagazine/.../N-almathm.pdf

9- الهوش، أبو بكر محمود. أدوات البحث عن المعلومات في البيئة الرقمية.ط.1. القاهرة: دار السحاب، 2013.

¹⁰ - إبراهيم، سيد ربيع سيد. الويب: نظام استرجاع المعلومات الكوني.المرجع السابق، ص47.

¹¹ - النشترتي، مؤمن صادق. التحديات التي تواجه خوارزميات محركات البحث في استرجاع المحتوى العربي دراسة مسحية تحليلية. المرجع السابق.

¹² - غيطاس، جمال محمد.فضاء إلكتروني جديد ومحركات بحث متعددة.[على الخط] زيارة يوم2016/10/04 متاحة في: <http://arab-librarians.Blospot.com>.

¹³ - إبراهيم، سيد ربيع سيد.نظم استرجاع قواعد بيانات الويب غير المرئية. الرياض: مكتبة الملك فهد الوطنية.2010. ص 174

¹⁴ Delengaigne, xavier.op.cit.p.10

¹⁵ - الإسكوا. تعزيز المحتوى الرقمي العربي في المنطقة العربية.[على الخط]. زيارة للموقع يوم 2017/08/08 متاحة على: <http://css.escwa.org.IB/ictd/1791/04a.pdf>

¹⁶ - غزال، عبد الرزاق. التواجد الإلكتروني باللغة العربية على شبكة الأنترنت: دراسة تقييمية لمواقع الويب الأكاديمية الجزائرية. رسالة دكتوراه. جامعة قسنطينة: قسم علم المكتبات، 2010. ص.53.

¹⁷ - المعتم، نبيل بن عبد الرحمان. البحث باللغة العربية على محرك البحث جوجل google مجلة مكتبة الملك فهد الوطنية.مج.17، ع.2، نوفمبر 2011.[على الخط] زيارة يوم 2016/04/12 متاحة على الموقع www.kfnl.org.sa/Ar/mediacenter/EMagazine/.../N-almathm.pdf

¹⁸ - الصويغ، علي بن سليمان. استرجاع المعلومات في اللغة العربية.ط.2. الرياض:مطبوعات مكتبة الملك فهد الوطنية، 2008. ص.15

¹⁹ - الدكتور، أيمن شعبان. المحتوى العربي على العنكبوتية العالمية: منظور معلوماتي- Cybrarians Journal - ع 23 سبتمبر 2010 [على الخط]. زيارة يوم 2015/12/04 متاحة على الموقع: http://journal.cybrarians.info/index.php?option=com_content&view=article&id=478:2011-08-12-09-15-31&catid=167:2009-05-20-10-03-11&Itemid=71

²⁰ - نبيل، علي. العرب وعصر المعلومات.عالم المعرفة.الكويت:المجلس الوطني للثقافة والفنون والآداب.1994.ص333.

²¹ - نشترتي، مؤمن سيد. التحديات التي تواجه خوارزميات محركات البحث في استرجاع المحتوى العربي دراسة مسحية تحليلية.المرجع السابق.

²² - المرجع نفسه.

²³ - نشترتي، مؤمن سيد. الشبكة العنكبوتية الدلالية : هوية تبحث عن الوجود : دراسة تأصيلية تحليلية. المرجع السابق.

10- ديور، عبد الرحمن غالب. نظم استرجاع المعلومات العربية واتجاهات
البحوث المعاصرة. مجلة إعلم. ع14. فبراير 2015 [على الخط] زيارة يوم 2016/01/08
متاحة على موقع <http://arab-afli.org/>

11- غزال، عبد الرزاق. التواجد الإلكتروني باللغة العربية على شبكة
الأنترنت: دراسة تقييمية لمواقع الويب الأكاديمية الجزائرية. رسالة دكتوراه. جامعة
قسنطينة: قسم علم المكتبات، 2010.

12- غيطاس، جمال محمد. فضاء إلكتروني جديد ومحركات بحث
متعددة. [على الخط] زيارة يوم 2016/10/04 متاحة في: <http://arab-librarians.Blospot.com>.

13- نبيل، علي. العرب وعصر المعلومات. عالم المعرفة. الكويت: المجلس
الوطني للثقافة والفنون والآداب. 1994.

14- نشرتي، مؤمن سيد. التحديات التي تواجه خوارزميات محركات البحث
في استرجاع المحتوى العربي دراسة مسحية تحليلية. - *Cybrarians Journal*. ع 29
(سبتمبر 2012) [على الخط] زيارة للموقع يوم 2015/10/15 متاحة
http://journal.cybrarians.info/index.php?option=com_content

15- نشرتي، مؤمن سيد. الشبكة العنكبوتية الدلالية: هوية تبحث عن
الوجود دراسة تأصيلية تحليلية. *cybrarians journal* ع 27. ديسمبر 2011 [على الخط] زيارة
يوم 2015/11/15 متاحة على الموقع <http://journal-cybrarians.info/index.php?option=com>

المراجع باللغة الفرنسية:

- Accart Jean.philippe;Rivier,Alescis. Mémento de l'information numérique.Paris:Editions du cercle de la librarié, 2012

- delengaigne,xavier. Organiser sa veille sur internet : au de la de google. paris :eyrolles;2012

denos, nathalie ; simili,karine. Organiser la recherche - 18
. Site visité [en ligne]d'information a l'ère numérique. Octobre 2011
le.13/11/2016. Disponible sur: c2i.univ-littoral.fr/ressources2011/fiche-d4-v1-0-20111020.pdf.

foenix-rio, beatrice. Recherche éveillée sur internet : mode d'emploi.paris :lavoisier;2011

- lardy, jean-pierre.la recherche d'information sur internet : 2001 outils et méthodes.7ed. paris : adbs;

- veronique, mesguich; armelle, thomas. Net 21 recherche2013.1ed. paris : abds;2013