

ندوة حول
"الرصد الاستراتيجي للمعلومات
الأساليب الحديثة للبحث عن المعلومات والويب الدلالي"

المحاضرون
الدكتور غسان مراد
أستاذ الألسنية المعلوماتية في الجامعة اللبنانية
الدكتور عماد بشير
أستاذ دراسات المعلومات وتكنولوجيا الإعلام
كلية الإعلام والتوثيق في الجامعة اللبنانية

مقدمة

ربما يكون الويب من أكثر الاكتشافات والتطورات إفساحا في المجال أمام حرية التعبير والنشر وتبادل المعلومات. ولكن، هناك مسؤولية على منتجي المعلومات وموفريها لنواحي رفع القيود وتذليل الصعوبات التي تؤدي إلى منع المستخدمين من الوصول إلى المعلومات.

ويعتبر الوصول إلى المواقع والبحث عن المعلومات من المحاور الساخنة والمهمة في مجال تطوير الويب، وفي البحث العلمي والأكاديمي. حيث بات يطرح حاليا مصطلح الويب الدلالي وأساليب أكثر دقة وعمقا في الوصول إلى المعلومات من خلال دلالات المفردات.

في هذا المجال، عقد المركز الاستشاري للدراسات والتوثيق - مديرية التوثيق ندوة عامة تحت عنوان الرصد الاستراتيجي للمعلومات الأساليب الحديثة للبحث عن المعلومات والويب الدلالي " أدار الندوة الأستاذ إبراهيم عيسى مدير التوثيق في المركز وحاضر فيها د. غسان مراد ود. عماد بشير. حيث تم الحديث عن الرصد الاستراتيجي واستعراض الطرق الدلالية للبحث عن المعلومات من خلال اللغة والصعوبات التي تعيق الوصول للمعلومات. كما تطرق إلى مراحل الويب 1 والويب 2 وصولا إلى الويب 3 أو الويب الدلالي، وكذلك مراحل تطور البحث عن المعلومات قبل وبعد الانترنت والتحديات التي تواجه الموثقين لناحية توثيق محتوى الوثائق أليا، والحديث عن أساليب البحث عن المعلومات. وسنتعرض في هذا البحث ابرز ما تم الحديث عنه في هذا السياق.

الطرق الدلالية للبحث عن المعلومات

الرصد الاستراتيجي للمعلومات:

الرصد الإستراتيجي (أو المراقبة) للمعلومات هو مصطلح أو عبارة تجمع عدة أنواع من الإستراتيجيات جيات يتخصص كل منها بمجال محدد ... منها مثلاً الرصد التكنولوجي، الرصد الاقتصادي، الرصد السياسي، والرصد القانوني، الخ...

إن تقنيات البحث الوثائقي وتقنيات معالجة المعلومات تساعد على اتخاذ القرارات الإستراتيجية المناسبة للمؤسسة أو للنظام أو للمنظمة.

من هنا فإن المصطلح "إستراتيجي" يستخدم ليشير إلى أنّ المعلومات المولدة من خلال الرصد الإستراتيجي لا تهتم بالعمليات الاعتيادية الجارية والمتكررة من وقت إلى آخر، بل تهتم أكثر بكل ما له علاقة بالمسائل الصعبة البنوية، والهيكلية، وقرارات غير اعتيادية.

إذاً، الرصد الاستراتيجي (أو المراقبة) للمعلومات هو مصطلح أو عبارة تجمع عدة أنواع من الاستراتيجيات أي التفتيش عن المعلومات. الهدف هو استباق الحدث بالتطور بمراقبة تغييرات تطورات السوق، التفاعلات التنافسية، تحليل المعلومات العلمية والتقنية. إنه دائم التطوير والتجديد. يهتم بالتفتيش عن المعلومات المرصدة كالإحصائية عبر حساب تواتر البيانات والمواضيع المطروحة، ومنها ما يركز على برمجة معلوماتية يهتم بها علمائون دون مراقبة للنصوص ومنها ما يركز على دراسة النصوص وعلوم الدلالة.

أما بالنسبة للموارد البشرية في زمن الويب، أصبح المسؤولون عن التوثيق داخل المؤسسات أو داخل مراكز التوظيف ذوي أهمية كبرى، إذ أنهم تحولوا إلى أشخاص يساهمون بأخذ القرارات، لا بل باتوا متخذي القرار الإستراتيجي للمؤسسة.

الأدوات المستخدمة للرصد تركز على محركات البحث، ولكنها مجدية فقط حسب الكشف الخاص لمجموعة معينة من المواقع. يوجد على الشبكة من ما هو مفهرس أي مرئي لمحركات البحث وهناك ما يسمى الويب المعقد أي غير المرئي من محركات البحث كبنوك المعلومات التي هي بحركية دائمة في

الزمن والمضمون والشبكات الخاصة الداخلية و انترانيت، والاكسترا نيت إلخ. والرصد يهتم بكل ما هو مفهرس، وبكل ما هو غير مفهرس. فالرصد الاجتماعي والسياسي يهتم بللتفتيش عن: "من قال"، "ماذا قال"، "كيف يقال"، "أين يقال" و"إلى من يقال". الجميع يراقب الكلّ والكلّ يراقب الجميع. فالمعلومات هي أساس كلّ القرارات مهما كانت طبيعتها وبالأخص ما يقال في الصحف ومواقع الويب . هذه البرامج لن تتم بشكل صحيح إلا من خلال المراقبة الألسنية للنصوص ووضع قواعد "ذكية" تسمح بالوصول إلى المعلومات وتخزينها لكي تستخدم لاحقاً.

الصعوبات في إيجاد المعلومات :

إن كثرة المعلومات تؤدي إلى صعوبة في إيجاد المعلومات الهادفة والمفيدة. حيث يكثر الشك في المعلومات المرصدة. إذا المراقب ومتابع المعلومات يقع دائماً تحت وطأة المعلومات المتجددة وكثرتها، وتحت وطأة الفوضى المعلوماتية التي تشكل الإنترنت حالياً المأوى المؤاتي لهكذا فوضى معلوماتية . وهذه الفوضى تؤدي إلى الخوف من عدم إيجاد المعلومة المناسبة من ناحية، ومن ناحية ثانية عليه إيجاد المعلومة في الوقت المناسب، لأن المعلومات أصبحت تعالج وتخزن وتجدد في وقت الزمن الواقعي. إن خلق برامج معلوماتية لمعالجة المعلومات (أدوات للرصد) هي من أهم المجالات حالياً على جميع الأصعدة. فهدف هذه الأدوات هو ترميز ، تصنيف، وتوسيم المعلومات بشكل مناسب يسمح بالدخول إلى المعلومات المناسبة لوجهة نظر المراقب وحاجته للمعلومات. كذلك أصبحت اللغات ذات أهمية كبرى للرصد الإستراتيجي وخاصة في السياق المعلوماتي العام . لهذا تلعب الألسنية دوراً مهماً لتحديد الاتجاهات الناجحة والفعالة في عملية "الفهم" المعلوماتي للنصوص المتعددة اللغات. ومن هنا كان تطوير برامج البحث عن المعلومات.

طرائق البحث الدلالي:

النصوص لها عدة دلالات وعدة أبعاد والهدف هو البحث الآلي عن أجزاء في النصوص . لذا يجب الاعتماد على بعض المؤشرات اللغوية لتساعد على تحديد أنماط النصوص. هذه المؤشرات توضع بشكل خوارزميات ضمن برنامج معلوماتي يقوم بالتجزئة الآلية.

إشكالية المعالجة الآلية:

بما أن اللغة هي الأداة الأولى للتعبير والتواصل وتمثيل المعلومات، تحتم وجوب معالجتها آلياً . هذا مبني على حقائق من أهمها الثورة الرقمية ، حيث إن جزء كبير من هذه الوثائق المنتجة بات موجوداً بشكل رقمي. إضافة إلى أن الترقيم قد أدى إلى ازدياد نسبة وجود المعلومات بكل أشكالها وأصبحت سلعة تباع وتشتري. بالتالي أصبح من الصعب على الإنسان أن يستهلك كل هذا الكم الهائل من المعلومات، ممّا استدعى استخدام الآلات أو البرامج التي تساعد على معالجتها. إذا أردنا من اللغة العربية أن تكون على المستوى المطلوب من مواكبة التطور التكنولوجي ، وجب معالجتها آلياً لكي ترقى وتتفاعل في إطار دينامية خاصة، إن كان في هيكليتها الداخلية أو في اثر ارتباط النظريات اللغوية الحديثة بها.

تعريف المعالجة الآلية للغة:

هي عملية خلق أنظمة وبرامج معلوماتية و "محطات عمل" تحلل النصوص المكتوبة والمنطوقة آلياً (أي بدون تدخل الإنسان) وذلك من أجل مساعدة مستعمل الحاسوب على حل بعض الأمور والمسائل المتعلقة باللغة؛ مثل برامج الترجمة الآلية، التصحيح الإملائي والقواعدي، الخلاصة الآلية، الإحصاء الآلي للمفردات، برامج إنتاج المصطلحات، الخ...

مراحل المعالجة الآلية للغة:

تمرّ المعالجة الآلية للغة بعدة مراحل، ولكل مرحلة مشاكلها وحلولها : كالتجزئة الآلية للنصوص، ومرحلة التحليل الصرفي والنحوي والدلالي ... الخ. هذه المراحل ببنيته تشبه الطبقات وكل طبقه تعتمد على سابقتها. (هذا بالنسبة ما يسمى المعالجة الكلاسيكية للغة، إذ يوجد حالياً طرق أخرى تعتمد على المدونات). إن المعالجة الآلية للغة تعتمد بشكل أساسي على الظواهر والقواعد اللغوية، فهي تتطلب دراسة ألسنية معمّقة لكل المراحل المذكورة آنفاً، إذ لا يمكن أن تتم هذه العملية إلا بالاعتماد على النظريات اللغوية المختلفة التي بدورها تساعد على رفع الالتباس وتحديد المعاني لكل جزء نصي حسب السياق اللغوي والمفردى والصرفي والنحوي والدلالي.

تبنى أنظمة المعالجة الآلية على التوافق بين الألسنية والمعلوماتية . الخطوات يقوم بها الالسنى المعلوماتي بالمراقبة الفعالة للنصوص، والتي ينجم عنها بناء قواعد تكتب بشكل عملائي، والنتيجة وضع هذه القواعد بشكل رمزي في نظام حاسوبي معيّن حسب البرنامج المطلوب . هذه الخطوات والمراحل هي شاملة في معالجة كل اللغات.

هل محركات البحث «ذكية»؟

إن محركات البحث مثل «غوغل» Google و «ياهو» Yahoo و «فوالا» Voila و «ألتافيسستا» Altavista وغيرها، يعتمد عملها عموماً على تقديم واجهة البحث عن النصوص باستعمال التقنيات الإحصائية، كأن تفتش النصوص بالاستناد إلى عدد الكلمات المطابقة لتلك المستخدمة في كتابة الطلب، ثم ترسل النصوص إلى المستخدم تراتبياً بحسب عدد تلك الكلمات داخل النص . وتشير الإحصاءات إن نسبة التغطية لمحركات البحث حالياً (إحصاءات أواسط ٢٠٠٩) لا يتعدى الـ ٢٠% من ما هو موجود على الويب..

هذه العملية سهلة نسبياً لكنها تقتصر إلى الدقة . فمن وجهة نظر علمية ولغوية، لا يحدّد العدّد أهمية الكلمات بالنسبة إلى النص . كما أن وجود كلمة في أحد النصوص، وإن تكرّرت، لا يعني أن الموضوع يتحدث عنها بالضرورة.

مثلاً وجود كلمة «فضاء» في نص لا يعني أنه يتناول الفضاء وعلومه . وفي المقابل، قد تحتوي بعض النصوص على معلومات مهمة عن الفضاء، من دون أن تكرر تلك الكلمة كثيراً في متنه . وهكذا، فأحياناً لا نحصل على المطلوب، وأحياناً أخرى نحصل على كمية هائلة من المعلومات والوثائق ليس في استطاعتنا قراءتها ولا مراجعتها . ويصف البعض هذا الوضع بعبارة «كثرة المعلومات تقتل المعرفة».

من ناحية ثانية، فإن لكل مستخدم هدفاً معيناً من المعلومات التي يريدتها في النص، وغالباً ما يتصل ذلك بمفهوم النص . فلنص يُفهم من كل شخص بحسب السياق . فلكل شخص تفسيره الخاص .مثلاً لا يبحث الصحافي في نص معين عن المعلومات عينها التي قد يسعى وراءها الباحث .ربما يهتم الصحافي بالبحث عن «ما قيل» و «من قال» حول موضوع ما، أما الباحث فقد يفتش عن العلاقات السببية والحلول المطروحة لإشكالية معينة داخل النص.

لهذا يجب التفكيك عن محركات بحث مختلفة ومتطورة، تعطي لكل حاجته من المعلومات التي يفترض وجودها ضمن نصوص تُنتقى بطريقة مناسبة أيضاً . ويتطلب هذا الأمر برامج معلوماتية تعالج النصوص بصورة مؤتمتة، أي أنها تتعامل مع المحتوى الدلالي للنص و تجزئه بحسب أنماط دلالاته . وفي هذه الحال، يصبح النص مجموعة من العلاقات بين مكوناته، والتي تنتظم بحسب المحتوى اللغوي لهذه الأجزاء . نجد في النص العناوين بأنواعها وترابطاتها، والخلاصات، والتعاريف، والعلاقات السببية، والجمل الأكثر أهمية التي يريد الكاتب أن يوصلها إلى القارئ، والاقتراب المباشر، والاقتراب المنقول، والعلاقات الزمانية والمكانية وغيرها . وتشكل هذه الأنماط الخريطة الدلالية للنصوص، خصوصاً تلك التي تتناول العلوم والإعلام.

الاستكشاف التناسي:

الطريقة التي نستعملها تعتمد على إحدى نظريات الأسنية الحديثة، وتسمى «الاستكشاف التناسي». هي من ضمن الأبحاث في مجال علوم المعرفة والتي تضع النقاط حول دور السياق في الفهم والتأويل . وترتكز على مبدأ أن السياق اللساني الذي يدخل مباشرة في تحديد المعاني وبنائها، يرفع التباس المفردات والقواعد والدلالات.

دور السياق في التفسير والتأويل :

إن الأبحاث في مجال علوم المعرفة تضع النقاط حول دور السياق في الفهم والتأويل، إن نظام اللغة المسنود من الذكاء الإصطناعي يعالج هذه المسألة باللجوء إلى ما يسمى معالجة عملية فهم القضايا . من ناحية أكثر أسنية نعتبر أن النصوص تحتوي على قدر كبير من المعلومات كافية إلى حد لا يستهان به لبناء المعنى بدون أن نحتاج إلى المعارف المربوطة بالمادة أو بالإختصاص أو الحالة المعالجة . بشكل عام نحن أمام عملية معرفية متصاعدة متعلقة بالمعلومات اللغوية تساعدنا على بناء معرفتنا على تنظيمها وعلى هيكلتها . السياق اللساني (اللغوي) يدخل مباشرة في تحديد وبناء المعاني . فهو يستخدم لرفع الإلتباس في المفردات . تعدد المعاني للمؤشرات اللغوية المتعلقة بالتطور ا لدياكرونيك للغة يساهم في إدخال التلاعب المعنوي واللجوء إلى السياق يساعد على تصحيح الإلتباس المعنوي لكي توصلنا اللغة إل ى الوظائف للتواصل ولتحديد العلاقات في الحوار .

كما إن تمثيل المعرفة مبني على بنية هيكلية تصاعدية متعلقة بعدة طبقات من التمثيل اللغوي وعلى عدة أنظمة تمثيلية : رمزية وتصويرية . هذا النظام الفكري التمثيلي للمعرفة يشبه النظام التنفيذي للبرامج المعلوماتية التي بحد ذاتها توصف العملية الفكرية للعقل الإنساني من خلال فهمه للإشارات . فالاستكشاف السياقي مبني حسب المؤشرات اللغوية الأساسية التي ترتبط دلاليًا بالفكرة أو بوجهة النظر المفتش عنها داخل النص وحسب المؤشرات اللغوية المكتملة التي تساعد على رفع الإلتباس وللتأكيد أن المؤشر اللغوي الأساسي له هذا المعنى في هذا السياق . المعاني هي متعددة للكلمات والنصوص . المبدأ الذي نعتمد عليه هو ترابط المعلومات . لم ندخل مرحلة الويب ٣ بعد، نعطي الشخص المعلومات المطلوبة دون الدخول إلى الانطولوجيا.

مراحل تطور الويب :

الويب ١ أو الويب الوصفي : يستخدم HTML التي تهتم بشكل النصوص .
الويب ٢ أو الويب الدلالي : ويطلق عليه (الويب التفاعلي) يستخدم XML لتوصيف النصوص بحسب المحتوى الدلالي.

الويب ٣ أو الويب الذكي : حيث تعمل محركات البحث كعميل ذكي تتعلم بذاتها تلقائياً من خلال تعامل المستهلك وسلوكياته. ويتحول النص المفرد إلى وحدة بإمكانها أن تتفاعل مع بقية النصوص.

مراحل البحث عن المعلومات :

أولاً- مرحلة ما قبل الانترنت:

توافرت المعلومات أمام الباحثين والمهتمين من خلال الإتصال المباشر قبل دخول إنترنت حيز الاستخدام الشعبي الواسع. حيث ظهرت في نهاية السبعينات من القرن الماضي خدمات المعلومات بالإتصال المباشر والفوري. كانت خدمة الوصول إلى قواعد معلومات مقننة تعتمد على التدخل الإنساني في التعبير عن المحتوى وعلى نظم بحث واسترجاع خاصة على مستوى المستخلص وبعض الإشارات البليوغرافية. لاحقاً، في منتصف التسعينات، بدأ توفير النصوص الكاملة وتعزيزت فكرة الإستغناء عن المتخصصين في تمثيل المحتوى. ترافق ذلك مع تطور هائل على مستوى كفاءات نظم الإسترجاع وإمكانات البحث عن مجموعة كلمات وعن جملة كاملة مع ربط بين الحقول المختلفة من عنوان وكاتب وتاريخ.... الخ. كذلك، تطور عمل نظم الاسترجاع إلى البحث في بداية النص، وآخره، أو فيه كاملاً مع تحديد عدد الكلمات المبحوثة وتكرارها في النص. ما زالت هذه الخدمات متوافرة في يومنا هذا مع كل إمكانات البحث المشار إليها:

Dialog, Lexis Nexis, Ebsco, Emerald, Science Direct...etc

عاد بعض هذه المواقع إلى توفير خيار البحث من خلال اللغة المقننة والإعتماد على المتخصصين للتعبير عن المحتوى، إضافة إلى البحث الحر في النصوص الكاملة. كما عاد الإعتماد على المكنز وما توفره من علاقات بين المصطلحات المفضلة وغير المفضلة. ولا يغيب أن الهدف الأساسي كان دائماً تقليل الإستدعاء المعني بعدد الوثائق المسترجعة من جهة وتكبير الدقة المعني بملائمة ما هو مسترجع من وثائق من البحث. كل هذه الخدمات متوافرة من خلال إنترنت وهي غير مرئية من محركات البحث لأنها مرسمة.

ثانياً - مرحلة الانترنت:

مع دخول الانترنت، أصبحنا أمام فضاء جديد تتوافر فيه مجموعة من المواقع الرقمية بتقنيات وتفصيل وشروط استخدام مختلفة يقسمها المتخصصون إلى رسمي وشبه رسمي في محاولة لتسهيل عملية التقييم لناحية المحتوى واعتماده في البحث العلمي أو الأخذ به كحقائق ومعلومات. العدد الأكبر من هذه المواقع لا يتيح الوصول المفتوح إلى محتواه. يستحيل الوصول إلى المعلومات التي تحتويها المواقع الإلكترونية المتوافرة على الانترنت دون الإعتماد على أدلة ومحركات البحث أو البوابات الرقمية، والحالات التي قد يصل فيها المستفيد إلى المواقع دون الإعتماد على أدوات البحث والتصفح تنحصر في معرفة عناوين هذه المواقع. حتى في هذه الحالات، سيكون صعباً البحث داخل تلك المواقع المعروفة إذا لم يكن فيها نظام بحث داخلي متطور.

تطور البحث في الانترنت:

بدأ مع أرشي وغوفر، بعدها فيرونيا وجاغهيد، ووندر، ثم ويب كراولر، ليكوس وإنفوسيك وأوبن تكست وياهو عام ١٩٩٤ وألتافيستا وإكسايت وهوت بوت. وفي ١٩٩٨ ظهر غوغل.

الفرق بين أدلة البحث ومحركات البحث عن المعلومات:

أدلة البحث	محركات البحث
<ul style="list-style-type: none"> ▪ يفضل استخدامها للبحث عن المواقع من خلال واصفات موضوعية وإسمية. ▪ تدخل إنساني مباشر في التعبير عن محتوى المواقع واختصاصاتها. ▪ تصنيف للمواقع المسجلة لديها حسب الموضوع ويقوم بذلك متخصصون في المجال. ▪ أكثر دقة وأقل استدعاءً لناحية الاسترجاع. ▪ قابلية التصفح خطوة خطوة من العام إلى الخاص حتى الوصول إلى قائمة المواقع المتاحة في الموضوع المتخصص. 	<ul style="list-style-type: none"> ▪ يفضل استخدامها للبحث عن الوثائق من خلال كلمات مفتاحية باللغة الطبيعية. ▪ اعتماد كلي على الآلة في كشف محتوى الوثائق التي تصل إليها. ▪ لا يعتمد تقنية تصنيف الوثائق حسب الموضوع. ▪ أكثر استدعاءً وأقل دقة لناحية الاسترجاع. ▪ لا مجال للتصفح الإنساني، بل زحف آلي للوصول إلى المعلومات.

أساليب البحث في أدلة ومحركات البحث:

- البحث البسيط
- البحث المتقدم مع أدوات الربط المعروفة: AND , OR, NOT وأحياناً NEAR
- البحث عن جملة، البحث المتداخل (Nesting)
- البتر
- تضيق نتائج البحث وتوسيعها
- تحديد من خلال اللغة، نوع الملف، البحث في موقع محدد، البحث في التاريخ لناحية حداثة الصفحة، مكان ورود الكلمات المفتاحية، تخصيص المنطقة، البحث الآمن، البحث في الصفحات المشابهة، البحث في الصفحات المربوطة بالعنوان المحدد،
- التعريف، أسماء الأفلام، المعادلات الرياضية، العملات، أرقام الهاتف...

الوصول إلى المعلومات:

الوصول إلى المحتوى المعلوماتي الخاص بموقع محدد (Accessibility) لا يعني بالضرورة الحصول على ما نريد من وثائق. قد يدخل هذا الأمر في مهارة تحديد المصدر. بينما يتطلب البحث عن الوثائق في المصدر المحدد مهارة من نوع آخر... إن سهولة الوصول إلى محتوى الموقع وتوفير البحث فيه والحصول على الوثائق المبتغاة يندرج في إطار ما هو معروف بـ "الإستخدامية" (Usability)، وهي من شروط الموقع الناجح.

الصعوبات والعيوب:

برزت صعوبات لناحية الوصول إلى المعلومات، المعوقات كثيرة بدءاً من التصفح مروراً باسترجاع اللمواد غير النصية والألو ان المستخدمة وشكل الحروف وعناصر كثيرة مرتبطة بالإستخدامية. أما عيوب البحث مرتبط بالإستدعاء الكبير للمعلومات ودقة قليلة.

بدأت محاولات التغلب على الصعوبات في البحث عن المعلومات مع الإنتقال إلى البحث في بيئة ويب ٣ الويب الدلالي.

المحتوى العربي على الانترنت:

على المستوى العربي الوجود الرقمي أقل من القليل بالنسبة لغيره من اللغات ضعيف ومفرداته وأسمائه مشتتة وفيها مشاكل جمة. ومحركات بحث وأدلة غير مكتملة وواهنة.

هناك ما هو أهم من الويب الدلالي : المبادرات المرتبطة بتعزيز المحتوى العربي على إنترنت، من منظمة الإسكوا، من السعودية ومن سورية