نشر خدمة البث الانتقائي للمعلومات عن طريق شبكة الانترنيت: نظام مقترح لجامعة الموصل

د. محمود صالح إسماعيل(*)

المقدمة

إن انفجار المعلومات الإلكترونية وتنوعها وضع تحديات كبيرة أمام نظم المعلومات الواسعة الانتشار في العالم. ففي مثل هذه البيئة الداينميكية الواسعة اصبح من الصعب على المستفيد التعامل مع البحث عن المعلومات عبر إجراءات محددة فقط كي يستطيع ملاحقة التطورات التي تحصل في مجال اختصاصه. فبدلاً من ملاحقة المستفيد للمعلومات، أصبحت المعلومات المختارة تصله حسب اهتماماته. لذلك دأبت الكثير من المكتبات ومراكز المعلومات على توفير مختلف خدمات المعلومات المنتقاة من خلال خدمة البث الانتقائي للمعلومات. حيث يقوم المستفيد بتثبيت اهتماماته الواسعة والعريضة، ثم يتم تقويم طلباته بشكل مستمر من خلال ملفات سمات المستفيدين (Profiles). وبناءً على ذلك سيستلم المستفيد وثائق ومصادر معلومات منتقاة استناداً إلى اهتماماته المثبتة في هذه الملفات. لذا اصبح

(*) أستاذ مساعد - قسم المعلومات والمكتبات - كلية الآداب / جامعة الموصل.

لخدمات البث الانتقائي للمعلومات أهمية كبيرة، وأصبحت هذه الخدمات من الوسائل التي لا يمكن الاستغناء عنها في نظم المعلومات العالمية.

لقد عمل الكثير من الباحثين على تسليط الضوء على توفير تنقية المعلومات المؤثرة والمقاسة بشكل جيد باستخدام مختلف الطرائق، فمنها ما اعتمد على العلاقات الموضوعية، ومنها ما اعتمد على قواعد استرجاع المعلومات الأساسية، ومنها ما اعتمد على الذكاء الاصطناعي[1]. ولم تتم دراسة فاعلية مثل هذه الطرائق إلا قليلا من خلال دراسة خدمة البث الانتقائي للمعلومات واستخداماتها في نظم المعلومات الكبيرة[2]. كما قام بعض الباحثين بدراسة استخدام كشافات ملفات سمات المستفيدين لتسليط الضوء على تنقية المعلومات في مركز المعلومات المركزي باستخدام تقنيات تنقية استرجاع المعلومات من اجل قياس الحاجة إلى استخدام خدمة البث الانتقائي للمعلومات وفاعليتها، التي عدّها هؤلاء الباحثين الخدمة المركزية[3].

إن عملية نشر البث الانتقائي للمعلومات تعاني من مشكلة المطابقة بين اهتمام المستفيد ومصادر المعلومات. فهناك موزعو المعلومات الذين يهتمون باهتمامات المستفيدين. يقابلهم المستفيدون الذين يبحثون عن المعلومات ذات العلاقة باهتماماتهم. لهذا يجب أن نبحث عن السبل والكيفية التي تضمن عملية المطابقة، بحيث تتوقف المعلومات من الموزعين إلى المستفيدين. والطريقة الوحيدة والبسيطة

(1) S. Loeb and D. Terry. Information filtering Communication of the ACM, 35 (12), 1992. pp. 26-81.

(2) T.W. Yan and H. Garcia-Molina. Index structure for information filtering under the vector space model. In: Proc. International Conference on Data Engineering, 1994. pp. 337-347.

(3) T.W. Yan and H. Garcia-Molina Index structure for selective dissemination of information under Boolean model. AGM transactions on Database Systems, 1994.

لعمل ذلك تكون عن طريق قيام المستفيدين بتثبيت اهتماماتهم على شكل ملفات سمات (Profiles) في مركز المعلومات. إن مثل هذه الخطة تبدو سهلة لكنها مكلفة إلى حد ما. لهذا يعصب أحيانا تحديد موقع مصادر المعلومات ذات العلاقة. ومما يعمق المشكلة قيام الموزعين بإرسال المعلومات إلى كل مستفيد لوحده، وينتج عن هذا استلام المستفيدين معلومات ليست ذات علاقة باهتماماتهم في بعض الأحيان. وهذا يؤدي إلى تضييع وقت شبكات المعلومات الدولية بنقل المعلومات التي ليس لها علاقة باهتمامات المستفيدين. والحل لمثل هذه المشكلة يكون عن طريق وجود وسيط ثالث هو خادمات البث الانتقائي للمعلومات (SDI Server) التي أصبحت معروفة من قبل الموزعين والمستفيدين عن طريق شبكة الانترنيت الدولية. والتي أخذت تتقبل ملفات سمات المستفيدين، وتجمع مصادر المعلومات، وتقوم بمطابقتها مع ملفات سمات المستفيدين. لهذا ستقوم هذه الدراسة بالإجابة عن التساؤلين الآتيين:

1. كيف يتم بناء ملفات سمات المستفيدين (Profiles)؟
2. كيف يتم توزيع مصادر المعلومات على خادمات البث الانتقائي للمعلومات؟

إن النظام المقترح في هذه الدراسة هو نظام مقترح لجامعة الموصل وقد تم تجريبه على عينة من المستفيدين من المكتبة المركزية لجامعة الموصل وعلى 740 دورية متاحة على الانترنيت من قبل الناشر البريطاني Black Wells مجاناً لمدة أربعة اشهر للتجريب. وقد توصل الباحث من خلال هذه التجربة إلى مجموعة من النتائج التي سيبينها في متن البحث.

شبكة الانترنيت والبث الانتقائي للمعلومات

إن دخول الإنترنت والانترانيت في حياة الناس، وتطورات شبكات المعلومات اللاسلكية المرتبطة بالأقمار الصناعية، فضلا عن توفر الاتصالات ذات الموجات العالية في مساكن المواطنين ساعد على تطوير مجالات واسعة من البث الانتقائي للمعلومات. وقد وفرت هذه المجالات الواسعة الكثير من البيانات والمعلومات للمستفيدين، والتي تتضمن مختلف أنواع وأشكال المعلومات والبيانات بما فيها من صحف ومجالات إلكترونية، ومعلومات ترفيهية، ومعلومات رياضية، ونظم معلومات المرور، فضلا عن المعلومات العلمية التي تصل إلى مكتب المستفيد أو منزله مباشرة. وامتداد هذه المجالات يعتمد أساسا على مجموعة حديثة من البيانات التي تأتى من مختلف مصادر المعلومات، ويتم تنقيتها استنادا إلى ملفات سمات المستفيدين التي تتضمن اهتماماتهم العلمية والشخصية، وبالتالي يتم إيصال هذه المعلومات إلى المستفيدين حسب اهتماماتهم أو اختصاصاتهم العلمية.

فمن اجل ضمان إيصال المعلومات المناسبة للمستفيد المناسب، تعتمد نظم البث الانتقائي للمعلومات على ملفات سمات المستفيدين. ونظم البث الانتقائي للمعلومات الجارية تستخدم واصفات بسيطة للمطابقة أو حزمة من الكلمات المفتاحية لاسترجاع المعلومات ولتقديم ملفات سمات المستفيدين ومطابقتها مع مصادر المعلومات الحديثة.

ومثل هذه الإجراءات تبدأ من إجراءات محددة لنقل اهتمامات المستفيدين لضمان عدم استلام المستفيد معلومات ليس لها علاقة باهتماماته العلمية. والعمل من خلال نماذج استرجاع المعلومات يعتمد على ملفات سمات المستفيدين اكثر من

اعتماده على التنقية الفاعلة. وفي محيط الانترنيت حيث تحتوي الشبكة على كميات كبيرة من البيانات المدخلة وعلى عدد كبير من المستفيدين الذين يستخدمونها، تكون الفاعلية والمعيارية هي الأساس[4].

وقد تم مؤخرا اعتبار لغة إكس إم إل (XML) بمثابة معيار للتبادل الآلي للمعلومات من خلال شبكة الانترنيت [5]. حيث إن هذه اللغة تسمح بترميز تركيبة المعلومات التي تتضمنها مصادر المعلومات. وهذه المعلومات يمكن تحويرها لبناء ملفات سمات المستفيدين الأكثر دقة والأكثر أهمية [6]. وبذلك ستكون هذه العملية مجدية والتي يتم من خلالها تسهيل مطابقة مصادر المعلومات مع ملفات سمات المستفيدين.

وقد ذكر مؤخراً أن شركة ما يكروسوفت تعمل على تصميم نظام تشغيل جديد يحل محل نظام التشغيل ويندوز XP وقد أطلق عليه اسم لونج هورن (Long Horn) وذكر أن هذا النظام سيعتمد على لغة حاسوبية مختلفة تماما هي لغة إكس إم إل (XML) . وقد ذكرت مايكروسوفت أن نظام لونج هورن يحتوي على واجهة تنظيم اتصال تقويم على أداء مهام متعددة، وهذا يعني تمكين مستخدم الحاسوب من القيام بعدة مهام، مثل كتابة رسالة عبر البريد الإلكتروني أو عمل صور رقمية عن طريق تلك الواجهة دونما حاجة إلى أساليب أخرى لاستكمال هذه

(4) Mehmet Altinel and Michael J. Franklin, Efficient filtering of XML. Document for selective Dissemination of information. In: Proceedings of the 26th VLDB Conference, Cairo, Egypt, 2000.

(5) T. Bray, J. Paoli, C. M. Sperberg-McQueen, Extensible Markup Language (XML) 1.0 http ://www.w3.org/TR/REC-xml.html, February, 1998.

(6) R. Cover, The SGML/XML Web Page, http ://www.oasis-open.org/cover/sgml-xml.html, December, 1999.

المهام كما يحدث حاليا مع ويندوز [7]. وهذا تأكيد لما ذهبنا إليه قليل بأن لغة إكس إم إل تعدّ معيارا للتبادل الآلي للمعلومات عبر شبكة الانترنيت.

وقد قام بعض الباحثين بتطوير نظام تنقية لمصادر المعلومات أطلق عليه منقي إكس (X Filter) حيث سيوفر هذا المنقي مطابقة عالية الكفاءة لمصادر المعلومات مع عدد كبير من ملفات سمات المستفيدين باستخدام لغة إكس إم إل. وباستخدام منفي إكس سيتم تقديم اهتمامات المستفيدين على شكل طلبات باستخدام لغة جديدة هي إكس باث (X Path). وجهاز منقي إكس يستخدم تركيبة تكشيف معقدة مع مجال لتعديل البيانات بسرعة واختيارها عبر ملفات سمات المستفيدين ذات العلاقة[8]. فنظام منفي إكس قد تم تصميمه بشكل يمكن العاملين عليه من توفير معيارية عالية. وهذا السبب وراء اختياره كتقنية مقنعة للعمل على نظم البث الانتقائي للمعلومات عبر شبكة الانترنيت.

استخدام لغة إكس إم إل XML لتصميم نظام البث الانتقائي للمعلومات:

إن خدمة البث الانتقائي للمعلومات تعتمد على عملية تنقية مصادر المعلومات المعتمدة على اهتمامات المستفيدين. والشكل(1) يرينا تصميم نظام البث الانتقائي للمعلومات المبني على لغة إكس إم إل، حيث هناك مجموعتان من المدخلات للنظام المتمثلة بملفات سمات المستفيدين (Profiles)، ومصادر المعلومات (الوثائق). فملفات سمات المستفيدين تتضمن المعلومات الخاصة

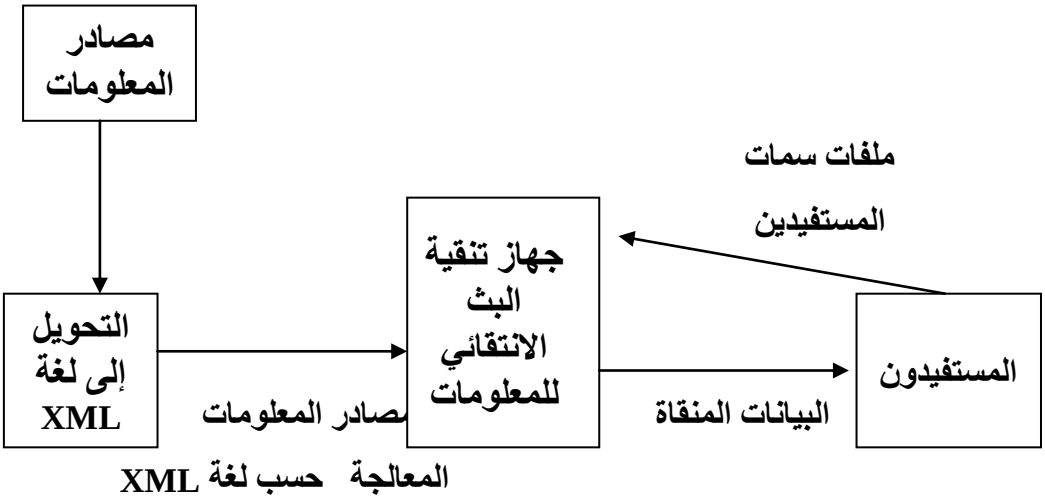
(7) لونج هورن يخلف ويندوز XP . في جريدة ومضات جامعية، ع 13، 1425هـ / تموز 2004م.

(8) J. Clark, S.DeRose, "XML Path Language (X Path) version 1.0." W3C Recommendation, http ://www.w3.org/TR/xpath; November, 1999.

باهتمامات المستفيدين. وفي معظم الأنظمة يتم بناء هذه الملفات من قبل المستفيدين أنفسهم، من خلال لوحة المفاتيح أو خلال الفأرة (Mouse) لإدخال البيانات النصية. وفي بعض الأنظمة يمكن تصميم هذه الملفات آلياً من خلال النظام عن طريق آلة لتلقيم المعلومات التي سيستخدمها المستفيدون. ثم يتم تحويلها إلى شكل يمكن خزنها بشكل آلي وتقويمها بكفاءة من قبل جهاز التنقية. وهذه الملفات تعدّ طلبات قائمة للمستفيدين تتعلق بكل مصادر المعلومات المدخلة إلى النظام.

والمجموعة الثانية من المدخلات في نظام البث الانتقائي للمعلومات هي مصادر المعلومات (الوثائق) التي سيتم تنقيتها. ويتركز كل الجهد على لغة إكس إم إل التي تقوم بترميز مصادر المعلومات. ولغة إكس إم إل هي لغة طبيعية لنظام البث الانتقائي للمعلومات لأنها أصبحت لغة آلية لإيصال المعلومات العلمية والإدارية الخاصة بالمنظمات والأفراد على شبكة الانترنيت. فضلا عن استخدامها في نشر المعلومات التجارية والمالية. حيث انها توفر آلية لتحديد محتويات مصادر المعلومات ووصف تنظيمها بشكل جيد. وتسمح بتنظيم مصادر المعلومات على شكل عناصر تحتوي على عناصر فرعية، وهذه العناصر يمكن تحديدها بعمق. فضلا عن العناصر الفرعية، حيث إن العناصر تتضمن بيانات على شكل نصوص أو عروض، ويمكن وضع قواعد عامة للعناصر والعروض التي يمكن تعريفها من خلال نظام تعريف نوع مصدر المعلومات، ونظام تعريف نوع مصدر المعلومات هذا يحدد تسمية هذه العناصر والعروض وطبيعة محتوياتها في مصدر المعلومات[9].

(9) S. DeRose, R Daniel Jr. E. Mater, "XML Pointer Language (X Pointer)", http ://www.w3.org/ TR/W Dxptr, December, 1999



الشكل (1) تصميم نظام البث الانتقائي للمعلومات المبني على لغة XML

إن نظام البث الانتقائي للمعلومات المصمم باستخدام لغة إكس إم إل يعتمد على جهاز التنقية. فحينما تصل المعلومات إلى جهاز التنقية، يتم مطابقته مع ملفات سمات المستفيدين لتحديد المستفيدين الذين يمكن إرساله إليهم. وفي حالة عمل نظام البث الانتقائي للمعلومات على شبكة الانترنيت، يتوقع أن يصل عدد المستفيدين من النظام إلى الملايين. والتحدي الكبير في مثل هذه الحالة هو البحث السريع والفاعل في هذا العدد الكبير من ملفات سمات المستفيدين لإيجاد مصادر المعلومات ذات العلاقة باهتماماتهم. ودور منقي إكس X Filter يكمن في حل هذه المعضلة.

إن الملف المستخدم في منقي إكس يعتمد على إكس باث X Path التي هي لغة لعنونة عناصر مصادر المعلومات في لغة إكس إم إل التي صممت للاستخدام في نقل المعلومات وترميزها، ولغة إكس باث توفر طريقة مرنة لتحديد المجالات، وتتعامل مع مصادر المعلومات وتتعامل مع مصادر المعلومات في لغة إكس إم إل

كشجرة من النقاط (Nodes). وهذه النقاط هي عبارة عن أنماط يمكن أن تطابق النقاط الموجودة في شجرة لغة إكس إم إل. وتقويم نمط لغة إكس باث يمكننا من تحديد العنصر الذي يمكن أن يكون على شكل حزمة من النقاط، أو أدوات الربط البولياني أو الأرقام، أو الخيوط[10].

والمجالات يمكن تحديدها كمجالات أساسية من جذور شجرة مصادر المعلومات أو كمجالات بديلة عن المواقع المحددة. ومجال الطلب يتضمن سلسلة من الخطوات أو قد يكون خطوة واحدة فقط. وفي هذه الخطوة يتم تحديد اسم النقطة (التي قد تكون اسم العنصر). كما يتم تحديد العلاقات الهرمية بين النقاط في الطلب باستخدام المعاملات الأبوية (" / ") ومعاملات جذر الكلمة السالفة (" / ") التي يتم أفرادها بأي رقم من هذه المستويات. فإذا كان الطلب يتضمن عناصر مصادر المعلومات السالفة لكل أجزاء المصدر، وفي هذه الحالة تعدّ عنصراً مباشراً من العناصر الأساسية لفهرس مصادر المعلومات. كما أن لغة إكس باث تسمح باستخدام معامل الحالة الشاذة أيضا (" * ") التي يمكن أن تطابق أي عنصر من الاسم في الطلب[11].

وفي منقي إكس يستخدم إكس باث لاختيار كل مصادر المعلومات بدلاً من أجزاء منها. وفي هذه الحالة تتعامل مع مجالات إكس باث كمستند عملي لمصادر المعلومات. فإذا طابقت مجالات إكس باث عنصراً واحداً على الأقل من مصدر المعلومات تعدّ مصدر المعلومات ذات مجال مقبول.

---

(10) J. Clark, "XSL transformations (XSLT) Version 1.0", http ://www.w3.org/TR/xslt, November, 1999.

(11) S. Derose. R. Daniel, OP. cit.

ويمكن استخدام إحدى لغات الطلب لتركيب البيانات بدلاً من استخدام إكس باث مثل لغة UNQI [12]، ولغة Lorel [13] ، ولغة XML-QL [14]. وقد اختار الباحث استخدام إكس باث في هذه الدراسة لسببين أساسين:

1. إننا نحتاج إلى لغة طلب كاملة لغرض تنقية مصادر المعلومات، خاصة وأن منقي إكس يختبر مصدر معلومات واحد في كل مرة، لهذا نحتاج إلى تحديد مجالات كل مصدر بشكل منفرد.
2. إن محددات إكس باث متبعة في الشبكة العنكبوتية العالمية، وهي لغة كفوءة في هذا المجال.

والإجراءات التي تم وصفها في هذه الدراسة يمكن أن تصلح لتحديد المجالات بشكل عام، وهذا يجعلها تصلح للتعامل مع لغات أخرى غير لغة إكس باث حسب الحاجة إلى ذلك.

جهاز التنقية Filter Engine

يتكون نظام البث الانتقائي للمعلومات المقترح من الأجزاء الأساسية الآتية:

1. معرف الكلمات الأساس لمصادر المعلومات المخزنة حسب لغة إكس إم إل.
2. معرفة لغة إكس باث لملفات سمات المستفيدين.

---

(12) P. Buneman, et. al. "A Query Language and Optimization Techniques for unstructured Data "Proc. ACM SIGMOD Conf., Montreal, Canada, 1996.

(13) S. Abiteboul, et. al. "The Lorel Query Language for Semi structured Data "International Journal on Digital Libraries, l (1). April, 1997. pp: 68-88.

(14) A. Deutsh, et. al. "XML-QL: A Query Language for XML" http :/www.w3.org/TR/NOTE-xml-al. August, 1998.

3. جهاز التنقية الذي يتولى عملية المطابقة بين مصادر المعلومات وملفات سمات المستفيدين.
4. القسم الخاص بالبث، الذي يقوم بإرسال المعلومات المنقاة إلى المستفيدين الحقيقيين.

فالجزء الأساس هو جهاز التنقية، الذي يستخدم كشاف معقد مع جهاز تعديل الموقع للتسريع في تحديد الموقع وتدقيق الملفات ذات العلاقة.

الخطوة الأولى في المطابقة تتم بعرض الملفات على مصادر المعلومات، ومن ثم يتم إرسال مصادر المعلومات إلى المستفيد الحقيقي. وآلية منقي إكس تستخدم نظام آلي لإيصال مصادر المعلومات المحدد إلى المستفيد صاحب الاهتمام المباشر. كما يمكن ربطه بنظام إيصال آلي لإيصال مصادر المعلومات آلياً إلى المستفيد عن طريق بريده الإلكتروني أو موقعه على شبكة الانترنيت.

إن جهاز التنقية الذي وصفت أجزاؤه في الشكل (2) يستلم الطلبات وكل طلب يحلل إلى مجموعة من الحالات من خلال معرفة إكس باث. وهذه الحالات تحدد النقاط في الطلب وتستخدم كحالات في جهاز التنقية، وهذه الحالات لا تشتمل الحالات الشاذة (" * ") وكل حالة تتضمن ما يأتي:

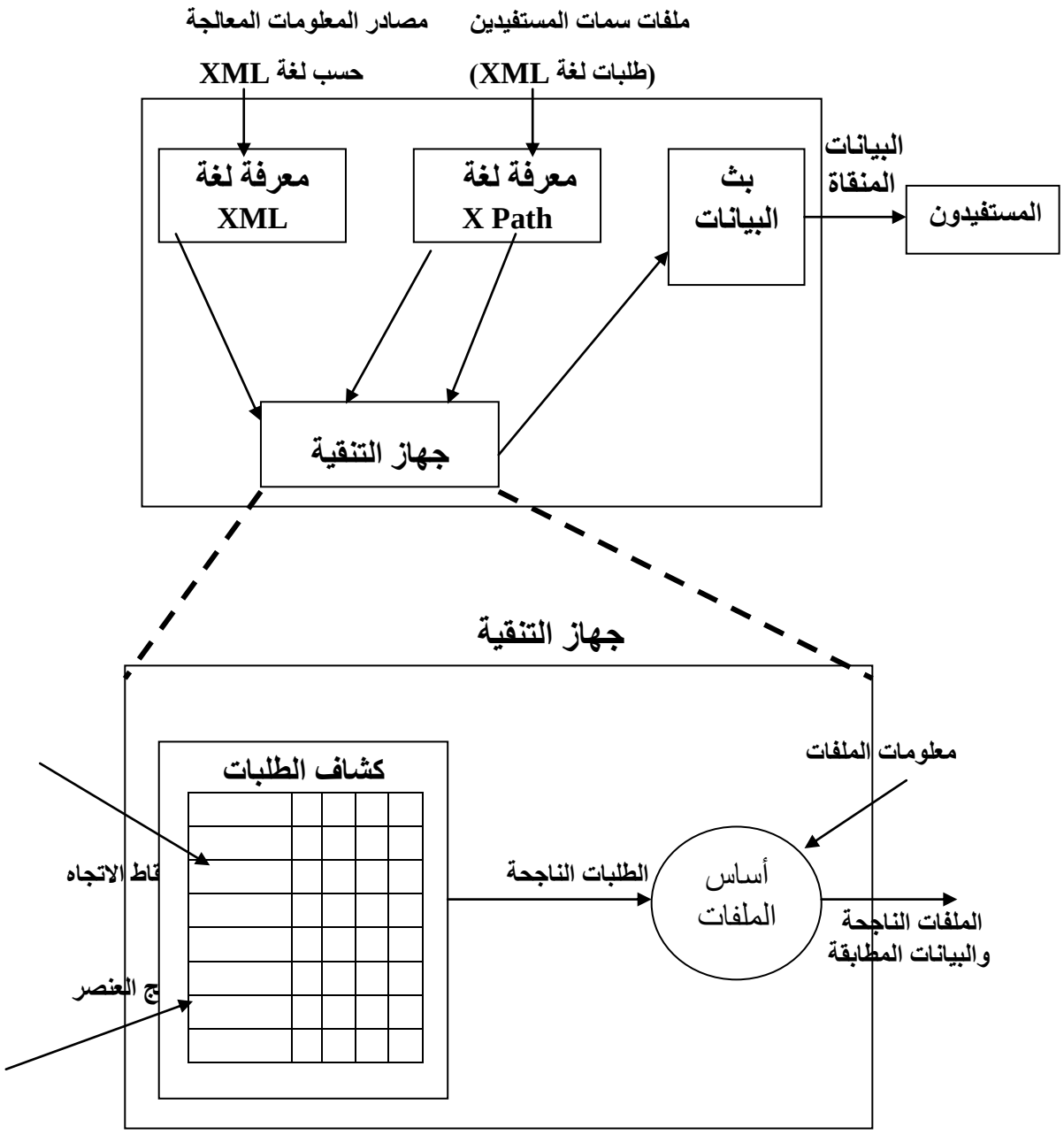
1. مميز الطلب (Query ld) : وهو عبارة عن مميز فريد لكل حالة ويبين انتمائها.
2. الموقع (Position): وهو رقم تسلسلي يحدد موقع كل حالة بين الحالات العديدة لكل طلب.

فالحالة الأولى تعطي الرقم واحد، والحالات اللاحقة ترقم ضمن التسلسل نفسه.

3. علاقة الموقع (Relative Pos.): وهو عدد صحيح يصف المسافة في مستويات مصدر المعلومات بين هذه الحالة والحالات السابقة. وهذه القيمة تعطي (صفر) لأول حالة إذا كانت لا تحتوي على معامل جذر الكلمة (" * "). والحالة التي تنفرد من الحالات السابقة من خلال معامل جذر الكلمة يتم تأشيرها بعلامة خاصة في موقعها وتكون قيمتها واحد. وفي الحالات الأخرى فإن علاقة الموقع للحالة هي مجموعة من واحد زائداً رقم الحالة الشاذة بين الحالات اللاحقة.

4. المستوى (Level): وهو عدد صحيح يوضح مستوى مصدر المعلومات في لغة إكس إم إل في المكان الذي يمكن فيه تدقيق الحالة. لأن لغة إكس إم إل لا تحصر أنواع العنصر عند ظهورها في مستويات مصدر المعلومات المتعددة، ولأن إكس باث تسمح للطلبات أن تكون محددة باستخدام العنونة، وليس من المحتمل ظهور هذه القيمة خلال تعريف الطلب. وهذه الحالة تختلف عن الحالات الثلاث السابقة إذ يمكن تحديث المعلومات خلال عملية تقويم الطلب.

فإذا كانت الحالة هي الحالة الأولى في الطلب وقد حدد لها مسافة واضحة في الأصل، يكون المستوى لهذه الحالة هو واحد زائداً المسافة الأصلية. أما إذا كانت قيمة علاقة الموقع للحالة هي واحد، ستكون قيمة مستواها واحد أيضا. وفي الحالات الأخرى تكون قيمة المستوى ضمن قائمة الصفر.



الشكل (2) معمارية منقي إكس X Filter

وبناءً على ذلك يتم تنظيم كشاف الطلبات كقائمة أساسية تعتمد على عناصر الأسماء التي تظهر في تعابير إكس باث بشكل أساس. ومع كل عنصر اسم فريد هناك قائمتان للحالات، الأولى قائمة الباحث، والثانية قائمة الانتظار.

فكل طلب حينما يكون منفرداً في استخدامه مع جهاز الحالة المحددة يكون محتوياً على نقطة مجال منفردة يتم تقديمها مع الحالة الجارية. وتسمى النقطة الجارية. فالنقطة الجارية لكل طلب يتم أدراجها في قائمة الباحث من خلال مدخل الكشاف وعلاقتها بعنصر الاسم. وكل نقاط المجال التي تتضمن الحالات المستقبلية تخزن في قائمة الانتظار من خلال علاقتها بعنصر الاسم. وفي حالة الانتقال إلى طلبات جهاز الحالات المحددة يتم تحديدها لتعزيز نقطة المجال بتحويلها من قائمة الانتظار إلى قائمة الباحث.

خادمات البث الانتقائي للمعلومات

إن تشغيل نظام البث الانتقائي المقترح يحتاج إلى عدد من الخادمات وطريقة العمل تكون بواحدة من الطرائق الثلاث الآتية:

1. إرسال مصادر المعلومات إلى أي خدام، وملفات سمات المستفيدين يتم تثبيتها في خدام واحد فقط. وفي هذه الحالة سيكون عدد الملفات في كل خدام قليلا إذا كان عدد الخادمات كبيراً، بينما سترتفع نسبة مصادر المعلومات التي ستصل إلى خدام البث الانتقائي من شبكة الانترنيت. فضلا عن أن توفر الخدمة سيكون قليلا، حيث إن الخدام ستقل خدمته وسيفقد المستفيدون الكثير من مصادر المعلومات.

2. تثبيت كل ملفات سمات المستفيدين في كل الخادمات، وترسل مصادر المعلومات إلى أي خدام منها. وفي هذه الحالة سيكون توفر الخدمة أعلى، حيث إن الخدام إذا انخفضت نسبة خدمته سيتم تمرير مصادر المعلومات إلى الخادمات الأخرى. وبذلك سيكون عدد ملفات سمات المستفيدين عال في كل خدام، وعملية تحديث هذه الملفات ستكون ذات كلفة عالية أيضا.
3. إن هذه الطريقة هي حل وسط بين الطريقتين السابقتين، وتكون عن طريق الإشارة إلى مجموعة الخادمات التي تحتوي على ملفات سمات المستفيدين بـ(س) ويتم إرسال هذه الملفات بواسطة س ع، ومجموعة الخادمات التي تحتوي على مصادر المعلومات يشار إليها بـ(ص) التي ترسل إليها مصادر المعلومات بواسطة ص غ ، ولكي تتأكد أن الملفات لا تفقد أي مصدر معلومات، يجب أن نضمن أن س ع لا تتقاطع مع ص غ لكل من س، ص.

وفي هذه الحالة سنتمكن من تحديد الاستجابات والخيارات التي نتبعها في التوزيع والتنظيم، ثم نتوصل إلى العلاقة بين قيم الاتصالات، وأوقات إيصال مصادر المعلومات، المستفيدون غالبا ما يتشابهون بالاهتمامات العامة. وفي هذه الحالة يمكننا تطوير أنموذج للسيطرة على التشابه بين الملفات. ويمكن لخدام البث الانتقائي للمعلومات الخاص بأخبار الشبكة أن يتعامل مع مثل هذه الحالات بكفاءة عالية وهذا يثبت ما ذهبنا إليه.

ولمقارنة مختلف التنظيمات ولتحديد أي من مجاميع الملفات اكثر أهمية، يمكن أن نستخدم تقويم الإنجاز التحليلي، الذي يستخدم تقنيات التنقية من اجل ثبات التقويم. وفي هذه الحالة يمكن استخدام طرائق استرجاع المعلومات

المعروفة مثل أنموذج فيكتر (VSM) المستخدم مع نظام وايس (WAIS) في خادمات البث الانتقائي للمعلومات[15]. حيث أن أنموذج فيكتر يقارن بين الملفات ومصادر المعلومات من خلال المصطلحات الكشفية، سواء أكان الملف على شكل رسومات، أم واجهة، أم ملف مستفيد. وتتم مطابقة مصدر المعلومات مع الملف إذا كان التشابه بين المصطلحات عاليا من خلال مصطلحات المستفيد التي حددها منذ البداية في ملفه. وخدام البث الانتقائي للمعلومات سيستخدم كشاف الملفات من خلال مصطلحات الملفات الكشفية لتسريع عملية المطابقة.
وهذه العمليات ضرورية لنمذجة عملية التنقية التي تتم في خدام البث الانتقائي للمعلومات[16].

وقد قام الباحث بتجربة هذا النظام على عينة مختارة من المستفيدين في جامعة الموصل بلغت (56) ملف سمات من خلال مطابقتها مع مقالات 740 دورية أتاحها الناشر البريطاني Black Wells مجاناً لمدة أربعة أشهر لجامعة الموصل. وقد تمت حوسبة القيم الأساسية باستخدام البيانات المعالجة في خدام البث الانتقائي لأخبار شبكة الانترنيت. والجدول (1) يعطينا بعض الإحصائيات لمدة أسبوع من 4 إلى 10 أيلول 2004. وهذه الإحصائيات تتضمن عدد الوثائق الموجود، وعدد ملفات السمات، وعدد التحديثات (وهي العمود 2، ،3، 4 على التوالي). وفي العمود (5) أعطينا القيمة النهائية للوظائف المطابقة

(15) G. Salton. Automatic Text Processing. Reading, Massachusetts; Addison Wesley, 1989.

(16) B. Kahle and A. Medlar. An Information system for corporate users: Wide Area Information Servers. Connections-The Interoperability Report, 5 (11), 1991, pp: 2-9.

(من ضمنها المكررة) وفي العمود (6) أعطينا العدد الصحيح للوثائق المطابقة. فالوثائق التي تطابق ملفين يتم حسابها مرتين في العمود (5) ومرة واحدة في العمود (6).

وقد حسبنا معدل الوثائق من خلال تقسيم عدد الوثائق الجديدة على عدد الثواني في اليوم. وتم حساب المعدل النهائي للوثائق المطابقة من خلال تقسيم معدل عدد الملفات على معدل عدد الوثائق الجديدة. وقد حسبنا معدل تحديث الملفات في اليوم من خلال تقسيم عدد الملفات على عدد الثواني في اليوم. وقد عددنا كل الملفات الموجودة على الخدام المحلي مجموعة واحدة. وفي المستقبل نتوقع زيادة موجات شبكة الانترنيت في جامعة الموصل، وفي الوقت نفسه نتوقع زيادة قيمة المعلومات، يصاحبها زيادة عدد السكان، وزيادة عدد مصادر المعلومات. ويعتقد الباحث أن الإحصائيات التي حصلنا عليها في الجدول المذكور آنفاً تؤشر إلى بداية مجموعة منسجمة من القيم لدراسة هذه الحالة.

ويمكن تقويم إنجاز مختلف الاستراتيجيات باستخدام ثلاثة مقاييس أساسية وهي:

مقياس تأخير إيصال مصدر المعلومات، وهو الذي يمثل الوقت المنقضي بين وقت إدخال مصدر المعلومات والوقت الذي يستلم فيه المستفيد مصدر المعلومات الذي يتلاءَم مع اهتماماته وهذا المقياس يتعلق بالطرائق التي تحسب وقت تقويم المعلومات، وتحديثها.

اليوم	عدد مصادر المعلومات المتوفرة	ملفات سمات المستفيدين	تحديث الملفات	المصادر المطابقة	
				العدد الكلي	العدد الحقيقي
9/4	49 ,649	56	41	20 ,717	10 ,504
9/5	43 ,034	54	43	19 ,703	9 ,956
9/6	26 ,874	55	49	16 ,340	7 ,690
9/7	30 ,645	52	45	16 ,436	8 ,172
9/8	41 ,003	54	38	22 ,629	12 ،664
9/9	61 ,523	56	40	20 ,474	9 ,981
9/10	38 ,967	55	47	19 ,493	9 ,726

الجدول (1) الإحصائيات المجمعة من خدام البث الانتقائي للمعلومات في المكتبة المركزية بجامعة الموصل

والمقياس الثاني هو مقياس الاستفادة من شبكة الانترنيت. وهو مقياس مهم في الحالات التي تكون فيها الشبكة مصدراً مهماً للمعلومات.

أما المقياس الثالث فهو مقياس توفر نظم نشر البث الانتقائي للمعلومات. وفي هذه الحالة نؤكد على فشل النظام نتيجة لضياع أو فقد مصادر المعلومات من المستفيد، وذلك في حالة عدم استلام المستفيد لمصدر المعلومات الذي يطابق اهتمامه بأسرع وقت ممكن.

إن هذه المقاييس تركز على حجم الوثيقة، وعدد ملفات سمات المستفيدين، ونسبة وصول مصدر المعلومات إلى المستفيد، ومدى مطابقة مصدر المعلومات مع ملف سمات المستفيد. وأخطر هذه الاعتبارات هو

حجم الوثيقة، الذي لا يؤثر فقط علي سير العمل في الشبكة، وإنما على وقت تنقية مصدر المعلومات.

فكلما نمت هذه القياسات طال وقت وصول مصدر المعلومات إلى المستفيد. فكيف نتعامل مع هكذا نظام؟ خاصة وإنا نتوقع زيادة كبيرة في عدد الوثائق، خاصة الوثائق المتعددة الوسائط (Multimedia) مثلا. فهل الحل يكمن في سرعة الشبكة؟ أم هل يكون الحل بزيادة خادمات البث الانتقائي للمعلومات[17]؟

وللإجابة عن هذه الأسئلة لا بد من القيام ببعض الإجراءات التي تساعدنا في السيطرة على تأخير وصول مصادر المعلومات، وهناك ثلاث حالات يمكن أن تساعدنا في حل هذه المشكلة وهي:

1. زيادة موجات الشبكة: وتتم بزيادة عدد الملفات وموجات الشبكة عشرة أضعاف ما تكون عليه في الحالات الاعتيادية. وينتج عن ذلك أنه إذا زادت تأثيرات موجات الشبكة يتم السيطرة على احتمالية زيادة مطابقة الملفات لمصادر المعلومات، وهذا يؤدي إلى زيادة نسبة وصول مصادر المعلومات للمستفيدين. فحينما يزداد عدد الملفات أو حجم مصادر المعلومات وحينما تكون نسبة وصول مصادر المعلومات عالية، تصبح خادمات البث الانتقائي للمعلومات عنق الزجاجة.
2. زيادة موجات الشبكة مع زيادة خادمات البث الانتقائي للمعلومات: وهذا يعني زيادة المقياس زيادة موجات الشبكة نفسها وعدد خادمات البث الانتقائي بشكل مناسب. حيث إن زيادة موجات الشبكة مع زيادة عدد الخادمات معاً

(17)T. W. Yan and H. Garcia-Molina Pp. Cit.

يمكننا من تحقيق إنتاج جيد على نطاق واسع. وهذا يعني أن النظام يمكن أن يتعامل مع عدد كبير من مصادر المعلومات وملفات سمات اكثر، اعتماداً على طول موجات الشبكة وعلى عدد الخادمات التي ستزداد بناءً على ذلك.

3. زيادة موجات مع تخفيض نسبة التنقية النظامية: وهذا يكون عن طريق زيادة موجات شبكة الانترنيت نسبياً، يصاحبها تقليص النسبة النظامية لوقت التنقية. حيث يمكن السيطرة على الوقت الذي يحتاجه خدام البث الانتقائي للمعلومات لتنقية مصدر المعلومات، لأن سرعة التنقية في خدام البث الانتقائي يكون تأثيرها أعلى من زيادة الخادمات للتعامل مع المقاييس ذات النسب الكبيرة باستخدام خطط التكشيف المقننة لتكشيف الملفات [18]، أو يكون ذلك عن طريق استخدام الحواسيب المتوازية لتقليص وقت المعالجة الذي يجب أن يؤخذ بالحسبان.

الاستنتاجات

لقد اقترحنا في هذه الدراسة نظام تنقية مصادر المعلومات المعتمدة على لغة إكس إم إل الذي سميناه منقي إكس للبث الانتقائي للمعلومات لاستخدامه في جامعة الموصل. ومنقي إكس يمكن المستفيدين من التعريف باهتماماتهم باستخدام لغة إكس باث للطلب. وهذه الطريقة تمكننا من بناء ملفات سمات للمستفيدين أكثر تعبيراً من نماذج ملفات السمات في نظم استرجاع المعلومات التي تساعدنا في استغلال تركيب المعلومات المتوفرة في مصادر المعلومات من خلال استخدام لغة إكس إم إل.

---

(18)T. W. Yan and H. Garcia-Molina Op. Cit.

ومنقي إكس هو الوسيلة الأساسية في مشروع نظام البث الانتقائي للمعلومات، الذي يمكن تطويره ليصبح أداة تتكيف مع مصادر المعلومات وتندمج مع آليات إيصال مصادر المعلومات المتعددة في بيئة مثل بيئة شبكة الانترنيت.

كما قمنا بوضع آلية لنشر خدمة البث الانتقائي للمعلومات، وقد وجدنا أن الخطة ملائمة لنشر مصادر المعلومات والملفات وهي الخطة التي تعمل موازنة بين حجم الملف ومصدر المعلومات، التي تكون أساسية في توفير خدمة فاعلة وكفوءة. إن آلية إيصال مصدر المعلومات إلى مجموعة الملفات اكثر فائدة من تقليص استخدام الشبكة وتأخير إيصال مصادر المعلومات. وللتدرج في القياس في حالة زيادة حجم الوثيقة، وزيادة نسبة المعلومات، وعدد الملفات، نحتاج إلى موجات ضخمة للشبكة، غير أن موجات شبكة ضخمة لوحدها لا تكفي، وإنما نحتاج إلى خادمات بث انتقائي للمعلومات اكثر تنظم ضمن خطة نشر جيدة، فضلاً عن ضرورة وجود نظم تكشيف كفوءة لمصادر المعلومات لكي تنجز عملية التنقية على خدام البث الانتقائي للمعلومات بكفاءة عالية.

Abstract

Distributed Selective Dissemination of Information through Internet: A proposal Project to Mosul University

Dr. Mahmood S. Ismail [*]

To help users cope with information overload, selective Dissemination of information (SDI) will increasingly become an important to olin wide area of information systems. In an SDI service, users post their long term queries, called profiles at some SDI servers and continuously receive new, filtered documents. To scale up with the volume of information and the size of users population, we need a distributed SDI service with multiple servers.

In this paper I decribe the advent of XML as a standard for information exchange and the developmeat of query languages for XML data enables the development of more sophisticated filtering mechanisms that take structure information in to account.

(*) Assitant Prof. – Department of Information and Librarianship – College of Arts / University of Mosul