مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (40) العدد (4) 2018
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (40) No. (4) 2018

دراسة مقارنة لخوارزميات الضغط وأثرها على تراسل المعطيات في الشبكات

دريد عجيب*

(تاريخ الإيداع 19 / 7 / 2018. قُبِل للنشر في 30/ 7 / 2018)

□ ملخّص □

نظرا للتزايد الكبير لاستخدام تراسل البيانات وخدمات تبادل المعلومات بأنواع مختلفة ضمن بيئات متباينة عتادياً وبرمجياً كان لابد من وجود لغة توصيف مثالية قابلة للتوسع والتطوير تخدم هذه الاحتياجات المتزايدة بأفضل شكل وبأقصر وقت ممكن و كانت اللغة الاكثر انتشاراً والأكثر استخداماً لغة XML.

كما أن اعتماد بنية الرسوميات احياناً خلق مشكلة أثرت على أداء شبكات نقل المعلومات نظرا للحجم الكبير للبيانات المتبادلة وكذلك الحاجة لسعة تخزينية كبيرة في طرفي الارسال والاستقبال

لذا كان لابد من إيجاد طرق فعالة لإنقاص حجم تلك البيانات التي يتم تبادلها من خلال الشبكة.

تم إجراء العديد من الأبحاث العلمية والتجارب العملية حول إيجاد طرق فعالة لإنقاص الحجم الفعلي للبيانات وباعتماد بارامترات مختلفة تؤثر على عملية ضغط الملفات بحيث تحقق نتائج أفضل بالتقليل من حجوم الملفات المتبادلة مع الانتباه الى أزمنة ضغط وفك الضغط للملفات.

لذا تم التركيز في هذا البحث على دراسة ومقارنة لبعض خوارزميات الضغط للملفات وبيان أثرها على تراسل البيانات في الشبكات.

الكلمات المفتاحية: لغة XML – بروتوكول SOAP– تقنية الضغط GZip– معدل الضغط CR– زمن الضغط CT معامل معدل الضغط Compression Ratio Factor (CRF)

* مشرف على الأعمال–الكهرباء–الكلية التطبيقية–جامعة تشرين–اللاذقية–سورية.

مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (40) العدد (4) 2018
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (40) No. (4) 2018

A comparative study of compression algorithms and their impact on data communication in networks

Doryd Ajeeb*

(Received 19 / 7 / 2018. Accepted 30 / 7 / 2018)

□ ABSTRACT □

Due to the large increase in the use of data communication and information exchange services of different types in different environments, the standard and the programming had to be a language of characterization is ideal for scalability and development that serve the growing needs in the best form and in the shortest possible time and was the most widely used language and the most widely used XML language.

he adoption of graphics architecture sometimes created a problem affecting the performance of information transmission networks due to the large volume of data exchanged as well as the need for large storage capacity at both ends of the transmission and reception.

Effective ways of reducing the amount of data exchanged through the network had to be found.

There have been many scientific researches and practical experiments on finding effective ways to reduce the actual size of the data and by adopting different parameters that affect the process of compressing the files so as to achieve better results by reducing the volumes of files exchanged with attention to times of compression and decompression of files.

In this research, we focused on the study and comparison of some compression algorithms for files and their effect on data communication in networks.

Keyword: XML Extensible Markup Language - Compression Ratio Factor CRF SOAP protocol Compression Time CT- Compression Ratio CR- - Gun Zip

* **Academic Supervisor, Department of Electrical, Faculty of Applied College, Tishreen University, Lattakia, Syria.**

مقدمة:

كان ظهور لغة XML وهي لغة نصية تستخدم لترميز البيانات بطريقة تشرح نفسها بنفسها وهو ما يعرف بالملفات ذاتية الوصف هذه اللغة كنتيجة للحاجة الماسة لوجود هيكلية ملفات تجمع بين قابلية تبادل المعلومات الواسعة التي توفرها الملفات النصية بالإضافة لفعالية وإمكانية التخزين القوية التي توفرها الملفات الثنائية ولذلك تم اعتمادها كخيار أمثلي لتبادل جميع أنواع البيانات عبر الشبكات وذلك بسبب بنيتها الهيكلية المساعدة على فهمها وتطبيقها في مختلف البيئات واللغات .

تم إجراء العديد من الأبحاث العلمية والتجارب العملية حول إيجاد طرق ضغط ملفات ذات أداء جيد لإنقاص الحجم الفعلي للبيانات التي يتم تبادلها من خلال الشبكة باعتماد بارامترات مختلفة تؤثر على عملية ضغط الملفات بحيث تحقق نتائج أفضل بالتقليل من حجوم الملفات المتبادلة مع الانتباه الى أزمنة ضغط و فك الضغط للملفات عند طرفي الارسال و الاستقبال .

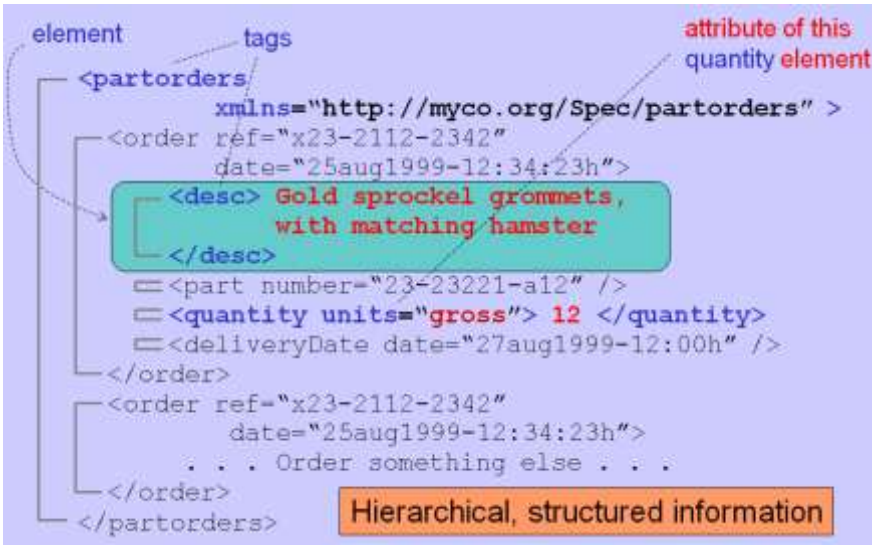
أهمية البحث وأهدافه :

نظرا لاستخدام لغة XML بشكل واسع في شبكات الأعمال لمكاملة التطبيقات ونقل المعطيات، لذا تم التركيز في هذا البحث على دراسة وايجاد أسهل وأفضل الطرق الموجودة لتحقيق طرق ضغط ملفات XML ذات أداء جيد. بحيث يمكن تحسين زمن إرسال ملفات XML عبر الشبكة وبالتالي تحسين أداء الشبكة ككل عن طريق إنقاص الحجم الفعلي لملفات XML وبالتالي إنقاص الزمن اللازم لإرسالها عبر الشبكة .

طرائق البحث ومواده

1- الشكل العام لملف XML (XML Format):

بالرغم من كل الميزات والفوائد التي يقدمها التمثيل بلغة XML إلا أن طبيعة ملف XML الغنية بالتراكيب والعناصر ينتج عن مشاكل من ناحية الحجم وكيفية استخدام تلك الملفات في بيئات آمنة الشكل (1) يوضح مفردات ملف ممثل بلغة XML :



الشكل (1) : الشكل العام لملف ممثل بلغة XML

- حيث يتألف ملف XML بشكل مبسط من مجموعة من العناصر elements مرتبة بشكل هرمي وفق محتوياتها تبدأ بعلامة دلالية tag مفتوحة وتنتهي بإخرى مغلقة تعبر عن محتوى هذا العنصر بالإضافة إلى عدد من الميزات attribute التي تخص كل عنصر على حدا ، وبهذا الشكل يمكن لملف xml أن يقرأ من قبل كل التطبيقات لأنه يكتب بطريقة تصف نفسها بنفسها ولكن هذا يكون على حساب الحجم الكبير الذي تشغله .
- تزايد استخدام XML في أغلب التطبيقات بما فيها تطبيقات الزمن الحقيقي. واحدة من أهم المشاكل المترافقة مع الـ XML هي النص الرسومي المستخدم والذي ينتج عنه أزمنة إرسال أكبر عبر الشبكة عدا الحاجة الى الكثير من مساحة التخزين عند الإرسال و الاستقبال.

2- الدراسات المرجعية :

لقد تم التطرق في العديد من الدراسات **المرجعية** الى خوارزميات ضغط ملفات XML

وأظهرت الدراسات قواعد الضغط وأدواتها كما أنها أظهرت عدة تقنيات ضغط منها Xmill و GZip وتم اجراء مقارنات بين تلك الخوارزميات من حيث معدل الضغط وزمن الضغط وفك الضغط ، كذلك بينت تلك الدراسة حركة مرور البيانات على الشبكة في كل خوارزمية ضغط. وكانت نتيجة هذه الدراسات بشكل عام تُبرز أنه وعلى سبيل المثال أن طريقة الضغط Xmill تحقق معدلات ضغط أعلى من طريقة الضغط GZip. كما ان طريقة الضغط GZip توفر معدل ضغط (40-55٪) كما أن طريقة الضغط Xmill توفر معدل ضغط يصل إلى 65 %

3 ضغط الملف (File Compression)[3]:

الطريقة الأكثر بساطة لتحسين تراسل ملفات XML عبر الشبكة هي إنقاص حجمها الفعلي بضغط الملف قبل إرساله. هذا يمكن أن يُنفذ بتقنيات ضغط موجودة مثل GZIP أو WinZip و WinRar أو ما شابه ذلك ،حيث أن الطبيعة النصية الرسومية لملفات XML تجعل استخدام الضغط مفيد جدا للحصول على نتائج جيدة من ناحية تقليل الحجم الفعلي للملفات.

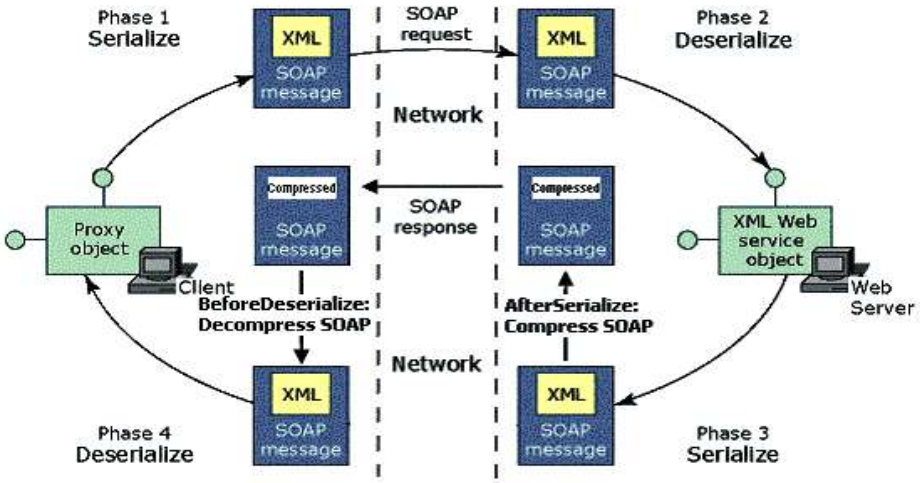
تم في هذا البحث دراسة و تحليل عمل و نتائج عدد من تقنيات الضغط لملفات XML مثل:

XComp & ICT Xpress & XGrind &Xmill

حيث لكل تقنية استخدامها الخاص حسب التطبيق المستخدم وأحجام الرسائل المرسلة.

زمن ضغط الملف قبل إرساله في الطرف المرسل وبطريقة مماثلة زمن فك ضغطه في الطرف المستقبل.تعد من أهم الأمور التي يجب العملة على تحقيقها ولاختصار الزمن اللازم لعمليات الضغط و فك الضغط تم اعتماد مبدأ ضغط رسائل الاستجابة (response) فقط نظرا لأنها تحمل كم كبير من المعطيات أما رسائل الطلب (request) فتكون كمية المعطيات فيها قليلة الحجم وممكن إرسالها بدون ضغط .

الشكل (2) يبين كيفية الضغط لرسائل الاستجابة على الطلب المرسلة عادة من المخدم إلى الزبون



الشكل (2) : عملية ضغط رسائل الاستجابة

تجدر الإشارة الى أن التعامل مع بروتوكول إرسال النصوص التشعبية HTTP (Hyper Text Transfer Protocol) و باستخدام بروتوكول SOAP حيث يتم استخدام تقنية ضغط ضمنية لبروتوكول HTTP وهي تقنية Gzip بشكل افتراضي في عملية الضغط. أما استخدام بروتوكول أخر للإرسال فان عملية الضغط ستكون من مهمة التطبيق المستخدم.

عملية الضغط وفك الضغط ليست عملية عديمة الكلفة لان الزمن اللازم للضغط وفكه يجب أن يؤخذ بعين الاعتبار حيث من الممكن أن يكون زمن إرسال الرسالة دون ضغط اقل من الزمن اللازم لإرساله مضغوطا في حالة الرسائل الصغيرة الحجم مثلا.

لذا لا بد من وجود خوارزميات تحكم ومراقبة لإجراء عملية الضغط في الوضع المناسب بحيث يؤدي إلى رفع كفاءة أداء الشبكة في كافة الظروف والحالات التي تطرأ على الشبكة.

4-تصنف طرق ضغط ملفات XML وفقاً لإدراكها لبنية وهيكلية الملف إلى [3]:

1. طريقة ضغط النص العام .... (General text compressors)
2. طريقة ضغط مدركة لملفات XML ....(XML-conscious compressors)

❖ طريقة ضغط النص العام:

تعد هذه الطريقة من أول الطرق التي استخدمت في عملية الضغط في هذا المجال ، سميت بهذا الاسم لأنها تقوم بتخزين الملفات بعد عملية الضغط كملفات نصية ، هذا النوع من طرق الضغط تتعامل مع ملف XML على انه ملف نصي بغض النظر عن محتوياته لذلك تعتبر من الطرق الغير مدركة لماهية ملف XML ، بعض الأمثلة لخوارزميات تعمل بهذه الطريقة من الضغط خوارزمية **GZip**

❖ طريقة ضغط مدركة لملفات XML

تختلف هذه الطريقة عن الطريقة الأولى فقد تم تصميمها للاستفادة من هيكلية ملف XML وذلك لتحقيق نسب ضغط أفضل من طريقة النص العام في الضغط ، بعض الأمثلة لخوارزميات تعمل بهذه الطريقة من الضغط خوارزمية **XMill** .

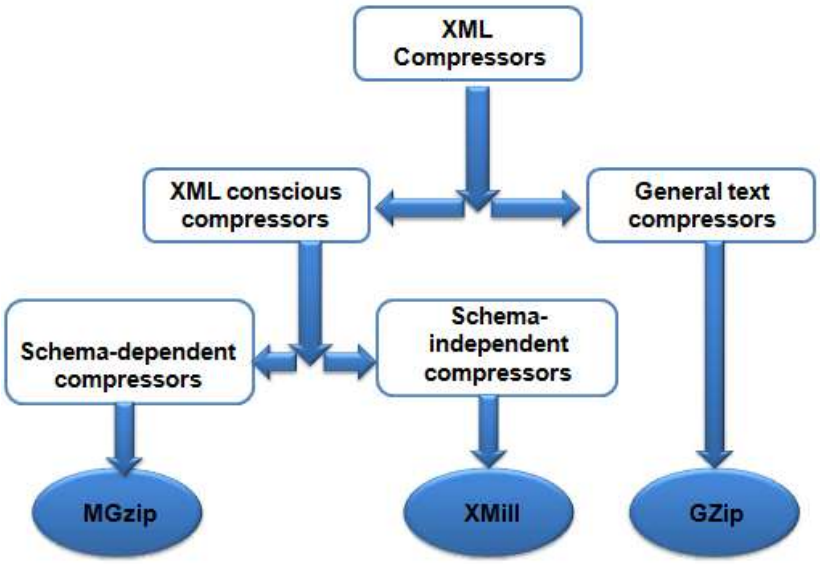
تقسم هذه الطريقة بدورها إلى مجموعتين وذلك بحسب توفر معلومات مخطط ملف XML:

1. طريقة ضغط مدركة تعتمد على مخطط المعلومات (معتمدة) Schema-dependent compression :

في هذه الطريقة لابد من الوصول لمعلومات مخطط ملف XML في كل من عملية الضغط وفك الضغط لتمام عملية الضغط ، كما في طريقة ضغط MGzip

2. طريقة ضغط مدركة لا تعتمد على معلومات مخطط XML(مستقلة) Schema-independent compressors :

لا يشترط توفر معلومات مخطط ملف XML لتحقيق عملية الضغط وفك الضغط كما في XMill. الشكل (3)



الشكل (3) : تصنيف طرق ضغط ملفات XML وفقاً لإدراكها لبنية وهيكلية الملف[3]

- **تصنف طرق ضغط ملفات XML وفقاً لدعمها لتنفيذ عملية الاستعلام [3]:**

1. طرق ضغط غير قابلة للاستعلام **(Non-queriable XML compression)**
2. طرق الضغط القابلة للاستعلام **(Queriable XML compression)**

❖ **طرق ضغط غير قابلة للاستعلام:**

لا تسمح هذه المجموعة لعملية الاستعلام على الملف المضغوط ،كما أنها تحقق نسب ضغط أفضل من مثيلاتها ، إن طريقة ضغط النص العام تنمي إلى هذا النوع من الطرق ، بعض الأمثلة عن هذه الطريقة **XMill** .

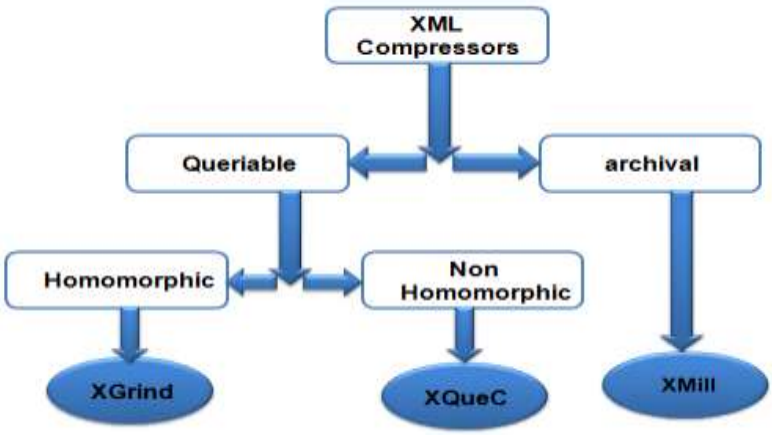
❖ **طرق الضغط القابلة للاستعلام:**

تسمح هذه المجموعة لعملية الاستعلام على الملف المضغوط كما أنها تحقق نسب ضغط أسوء من مثيلاتها، لها القدرة على تنفيذ استعلامات مباشرة على ملف XMLمضغوط ، هذه الميزة مهمة للعديد من التطبيقات التي يتم استضافتها على أجهزة الكمبيوتر ذات الموارد المحدودة، مثل الأجهزة المحمولة وأنظمة تحديد المواقع ، تندرج هذه المجموعة تحت تصنيف طرق الضغط المدركة لملفات XML.

تقسم طرق الضغط القابلة للاستعلام إلى:

1. **طرق ضغط متماثلة:** حيث يتم الاحتفاظ ببنية ملف XML الأصلي ، إن بنية الملف الأصلي والمضغوط تكون متشابهة ' من الأمثلة على هذه الطريقة XGrind.

2. **طرق ضغط غير متماثلة:** إن بنية الملف الأصلي والمضغوط تكون متشابهة ' من الأمثلة على هذه الطريقة XQueC



الشكل (4) : تصنف طرق ضغط ملفات XML وفقاً لدعمها لتنفيذ عملية الاستعلام[3]

5-مزايا استخدام تقنية الضغط لملفات XML :

- تقليل الزمن اللازم نقل بيانات XML عبر الشبكة.
- معالجة مشكلة حجم ملفات XML .
- تقليل عرض النطاق الترددي للشبكة المستخدم لتبادل البيانات هذه.
- تقليل مساحة القرص المطلوبة في التخزين.
- التقليل من متطلبات الذاكرة الرئيسية التي تستخدم عند معالجة واستعلام ملف XML.

ويعتبر تقييم خوارزميات الضغط على بارامترات أساسية مثل :

- زمن الضغط وفك الضغط ..... Decompression & Compression Time
- معدل الضغط ..... Compression Ratio
- الملف المراد ضغطه ..... plain text XML
- حجم الملف بعد الضغط ..... Decompression text XML

6 ميزات خوارزميات الضغط

GNU zip (GZip) : [9]

تستخدم هذه الطريقة في ضغط صفحات الويب لتخفيف الضغط على متصفحات الانترنت،،صممت هذه الطريقة بالاعتماد على HTTP وخوادم الويب كميزة قياسية. وهي خوارزمية مفتوحة المصدر تستخدم للأغراض العامة و توفر

معدل ضغط (40–50٪) ولا تتطلب أي معرفة مسبقة ببنية الملف المراد ضغطه ولكن من مساوئها أن ضغط العناصر فيها قد يكون محدد بسبب الاعتماد على العناصر والسمات بعيدة المدى

Xmill [1] :

لا تحتاج لمعرفة نوع وبنية الوثيقة المراد ضغطها وتحافظ على ملف الإدخال XML

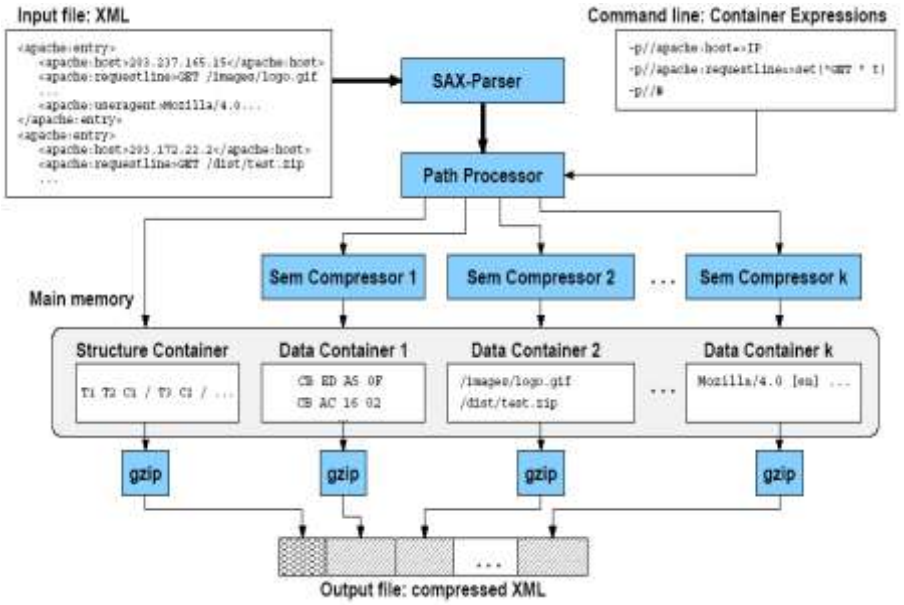
وتحقق معدلات ضغط أفضل من طرق ضغط مثل GZIP وقد تصل إلى الضعف

و من مبادئ الضغط في Xmill

- استخدام هيكلية لفصل البيانات بمجموعات.
- تجميع البيانات المتعلقة بعضها ببعض في حاوية واحدة وضغط كل حاوية على حدا.
- تطبيق طريقة الضغط المناسبة لكل حاوية بما يتناسب معها .

بنية ال XMILL :

- SAX Parser : الذي يضع الرموز token على مسار المعالج للمعالجة .
- Path Processor : يقوم بتحديد الحاويات وانتماء كل مجموعة لحاوية ما .
- Semantic Compressors : يقوم بضغط كل حاوية على حدا مع اخذ نسخة من الحاوية قبل ضغطها وتخزينها في الذاكرة .
- عند امتلاء النافذة المحددة تبدأ عملية الضغط باستخدام خوارزمية GZip الشكل(4)



الشكل (4) : بنية العمل لتقنية الضغط XMILL [1]

تحقق هذه الطريقة أفضل معدل ضغط مقارنة بالطرق الأخرى والاهم أنه ليس على حساب سرعة العمل، وذلك نتيجة فصل البنية عن المحتوى بالدرجة الأولى. وهذا أمر مهم في عملية أرشفة البيانات وبالتالي ستتطلب مساحة أقل على القرص.

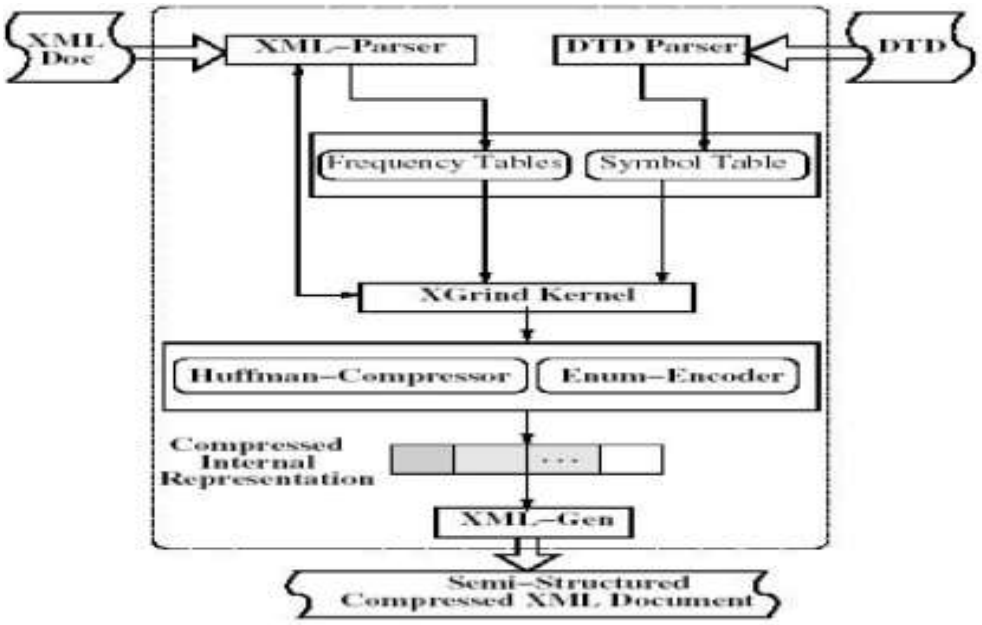
و لكن من عيوب هذه الطريقة أنه لا يمكن استخدام الملف الناتج عن عملية الضغط أو الاستعلام عنه حتى يتم فك الضغط (XdMILL)

هذه الخوارزمية تعطي نتيجة خوارزمية GZIP ذاتها في حال حجم الملف المدخل (المراد تشفيره) أقل من 20 KB وتستخدم هذه الطريقة لضغط البيانات والنصوص وليس لضغط الصور والفيديو.

XGrind [4]:

هي طريقة ضغط لمستندات XML وتتم عملية الضغط للوثيقة وفق هذه التقنية وبالوقت نفسه يتم الاحتفاظ ببنية الوثيقة الأصلية هذا يسهل من عملية معالجة الملف المضغوط.

هذه الميزة مهمة جداً بالنسبة لأجهزة الحواسيب ذات الموارد المحدودة وبالتالي التقليل من زمن الانتظار وسرعة في الاستعلام الشكل (5).



الشكل (5) : بنية العمل لتقنية الضغط XGrind [4]

تحقق هذه التقنية ما يسمى بالضغط المتماثل homomorphic compression

والذي يتميز بتقنيات تحليل الكفاءة والقدرة على بناء مؤشرات للوثيقة المضغوطة نفسها بنفسها وقابلية تطبيق التحديثات على الملف المضغوط وإمكانية التأكد من صحة الملف المضغوط.

كما تحقق طريقة XGRIND معدلات ضغط أقل من طريقة XMILL

وتظهر تحسينات كبيرة في وقت الاستجابة والاستعلام مع زيادة استخدام عرض النطاق الترددي بزيادة كمية البيانات. ولكن لا تدعم عدة عمليات مثل عمليات الربط، والتجميع، والاستعلام ات المتداخلة أو بناء العمليات وتستخدم هذه الطريقة إستراتيجية الاتجاه من الجذر نحو الأوراق **XComp : [9]**

هذه التقنية تشابه الـXMILL في الضغط حيث يتم فصل الهيكلية وترميزها كسلسلة من الأعداد الصحيحة، بينما يتم تجميع البيانات عن العلامات في شجرة الوثيقة

العلامات، والصفات، وعناصر البيانات تلعب دورا رئيسيا في تطوير XComp وتستخدم هذه الميزة بشكل رئيسي من أجل استخدام أساليب محددة لمعالجة الهيكلة بحيث يمكن تحسين نسبة الضغط بشكل كبير. سيتم ضغط البيانات المتعلقة معا بعد إعادة التنظيم.

تعتمد آلية الضغط هنا على أنه يتم التعامل مع مختلف عناصر وثيقة XMLالمدخلة بشكل منفصل في الضغط. ولا يتم تحليلها بل يتم تخزينه كما هي كبيانات و من ثم يتم إعادة تنظيم البيانات المرتبطة ببعضها ليتم ضغطها معا

النتائج والمناقشة:

تم في البحث مقارنة خوارزميات الضغط المذكورة سابقاً بعد تطبيقها على عينات عشوائية من ملفات xml من حيث حجم الملف الأصلي والمضغوط – زمن الضغط وفك الضغط – معدل الضغط – زمن الإرسال للملف المضغوط والملف الأصلي واستخلاص النتائج. باعتماد العلاقات التالية:

- زمن الضغط ..... Compression Time (CT)
- معدل الضغط .....Compression Time (CR)

$$CR = 1 - \frac{\text{Compressed Size}}{\text{Original Size}} * 100 \quad [8,11]$$

- معامل معدل الضغط .....Compression Ratio Factor (CRF)

$$CRF = \frac{CRxGrind}{CRxMill} \quad [8]$$

- معامل زمن الضغط .....Compression Time Factor (CTF)

$$CTF = \frac{CTxGrind}{CTxMill}$$

تمت المحاكاة على مجموعة عينات عشوائية من ملفات الـ XML و بحجوم مختلفة كما هو مبين في الجدول 1:

الجدول(1) مجموعة عينات عشوائية من ملفات الـ XML و بحجوم مختلفة

رقم الملف	حجم الملف in Mbytes	ملفXML
1	1100	xmlbenchmark
2	390	conferences
3	290	journals
4	165	shakespeare
5	365	xpress
6	960	student1
7	1360	student4

الجدول (1) عينة عشوائية من ملفات XML

مقارنة حجم الملف الأصلي المضغوط والفرق ...في كل من XGrind و Xmill :

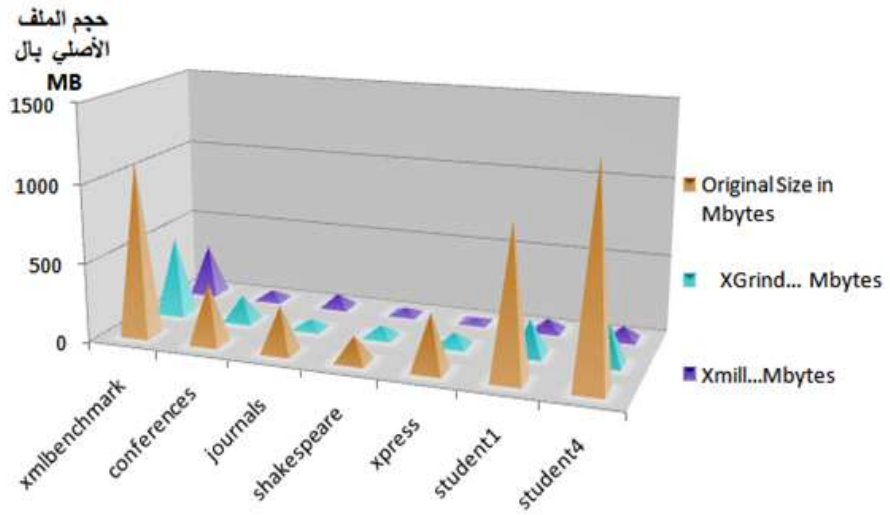
تم تطبيق كل من تقنيتي Xmill و XGrind على العينات السابقة وحساب الفرق بين حجم الملف قبل الضغط وبعده لكل تقنية.

الجدول (2) نتائج تطبيق تقنيتي XGrind و Xmill

الملف	حجم الملف الأساسي in Mbytes	حجم الملف باستخدام تقنية الضغط XGrindMbytes	الفرق بين حجم الملف قبل الضغط و بعده	حجم الملف باستخدام تقنية الضغط Xmill Mbytes	الفرق بين حجم الملف قبل الضغط و بعده
xmlbenchmark	1100	503.55	615.45	325.06	793.94
conferences	390	162.58	219.42	58.828	323.172
journals	290	57.85	236.15	85.59	208.41
shakespeare	165	72.45	88.55	41.86	119.14
xpress	365	83.03	277.97	25.27	335.73
student1	960	220	740	76.8	883.2
student4	1360	247.5	1127.5	82.5	1292.5

الشكل (5) يبين مقارنة بين تقنيتي XGrind و XmillXGrind و Xmill

- نلاحظ من النتائج أن حجم الملفات المضغوطة بتقنية XGrind أكبر من مثيلتها في XMill . وبالتالي إن تقنية XMill أفضل من XGrind من حيث حجم الملفات المضغوطة.



الشكل (5) : مقارنة بين نتائج تقنيتي XGrind و Xmill

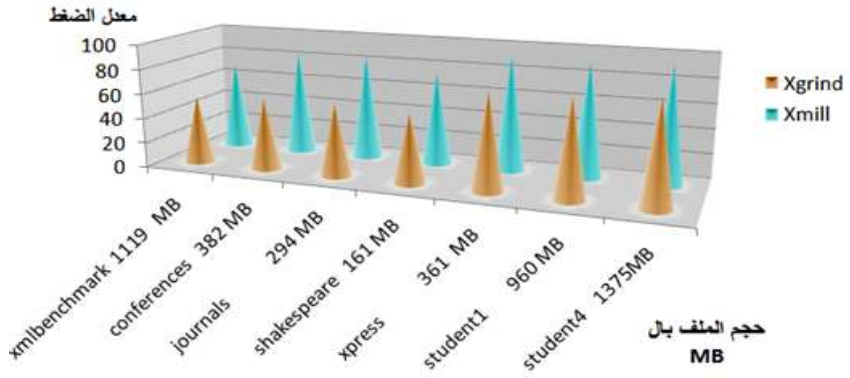
مقارنة معدل الضغط لكل من XGrind و Xmill :

يبين الجدول (3) مقارنة بين نتائج تقنيتي XGrind و Xmill بالنسبة لمعدل الضغط حيث تم تطبيق كل من تقنيتي Xmill و XGrind على العينات السابقة وحساب معدل الضغط CR.

الجدول (3) مقارنة بين نتائج تقنيتي XGrind و Xmill بالنسبة لمعدل الضغط

الملف	حجم الملف الأساسي in Mbytes	معدل الضغط CR Xmill…%	معدل الضغط CR XGrind…%
xmlbenchmark	1100	70.95	55.03
conferences	390	84.61	57.44
journals	290	85.59	57.85
shakespeare	165	74.12	54.96
xpress	365	93.54	76.85
student1	960	91.74	77.13
student4	1360	93.87	82.12

نلاحظ من النتائج أن أن معدل الضغط في XGrind أقل من مثيلتها في .Xmillوبالتالي فإن تقنية XMill أفضل من XGrind من حيث معدل الضغط للملفات المضغوطة الشكل (6)



الشكل (6) : مقارنة معدل الضغط لكل من XGrind و Xmill

مقارنة زمن الضغط المستغرق بالثانية لكل من XGrind و Xmill :

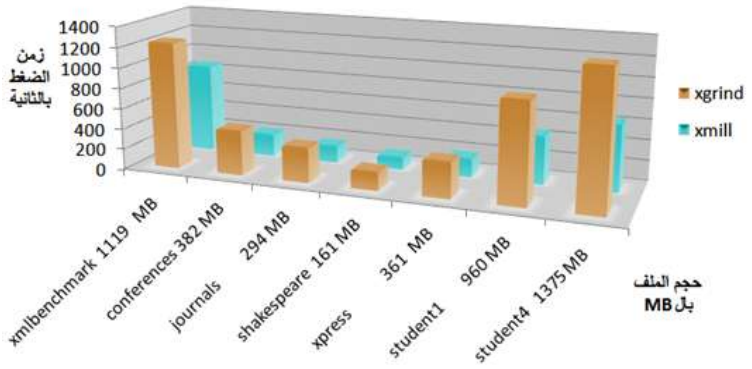
تم تطبيق كل من تقنيتي Xmill و XGrind على العينات السابقة وحساب زمن الضغط بالثانية لكل تقنية CT.

و يبين الجدول (7) نتائج تطبيق تقنيتي XGrind و Xmill والفرق بينها بالنسبة لزمن الضغطCT

الجدول (7) نتائج تطبيق تقنيتي XGrind و Xmill والفرق بينها بالنسبة لزمن الضغطCT

الملف	حجم الملف الأساسي in Mbytes	زمن الضغطCT Xmill sec	زمن الضغطCT XGrind sec
xmlbenchmark	1100	878	1246
conferences	390	222	442
journals	290	170	344
shakespeare	165	125	183

xpress	365	182	353
Student1	960	471	978
student4	1360	647	1328



الشكل (7) : مقارنة زمن الضغط المستغرق بالثانية لكل من XGrind و Xmill

- النتائج الرقمية والمخططات البيانية تبين أن زمن الضغط في XGrind يساوي تقريباً ضعفي الزمن في مثيلتها Xmillوبالتالي فإن تقنية XMill أفضل من تقنية XGrind من حيث زمن الضغط للملفات المضغوطة

ليكن لدينا كما هو مبين في الجدول (8) عينات عشوائية من ملفات XML بحجوم مختلفة صغيرة نوعاً ما

الجدول (8) عينات عشوائية من ملفات XML

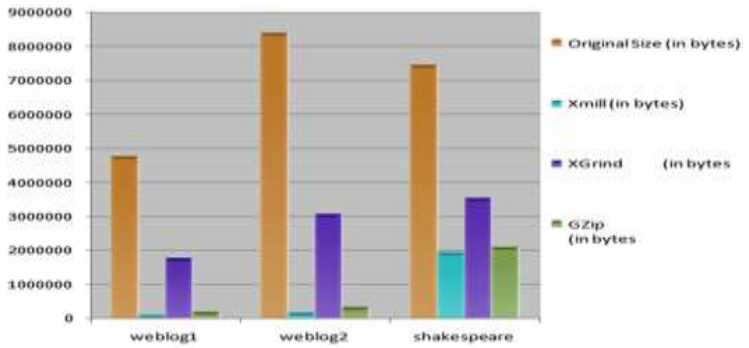
الرقم	حجم الملف الأساسي (in bytes)	الملف
1	4746593	weblog1
2	8387865	weblog2
3	9434567	shakespeare

مقارنة حجم الملف الأصلي والمضغوط لكل من XGrind و Xmill و GZip :

تم تطبيق تقنيات الضغط التالية Xmill و XGrind و GZip على العينات الثلاث السابقة وتم حساب الفرق بين حجم الملف المضغوط و الملف الأصلي وتسميته كما هو واضح في الجدول (9)و **الشكل (8)**.

الجدول (9) نتائج تطبيق تقنيتي XGrind و Xmill و GZip

حجم الملف بعد الضغط GZip(in bytes)	حجم الملف بعد الضغط XGrind (in bytes)	حجم الملف بعد الضغط Xmill (in bytes)	حجم الملف الأصلي (in bytes)	الملف
198021	1754476	93966	4746593	weblog1
317602	3072274	157524	8387865	weblog2
2084979	3544311	1935738	9434567	shakespeare



الشكل (8) يبين مقارنة لحجوم الملفات الأصلية مع حجومها بعد الضغط لكل من XGrind و Xmill و GZip

- المخططات البيانية تظهر أن حجم الملفات المضغوطة بتقنيتي XGrind و GZip أكبر من مثيلتها في XMill وبالتالي إن أفضل طريقة للضغط وفق المعطيات المدروسة هنا هيXmill وتأتي بعدها GZip ثم XGrind.

من ناحية مقارنة زمن الضغط وفك الضغط لكل من XGrind و Xmill:

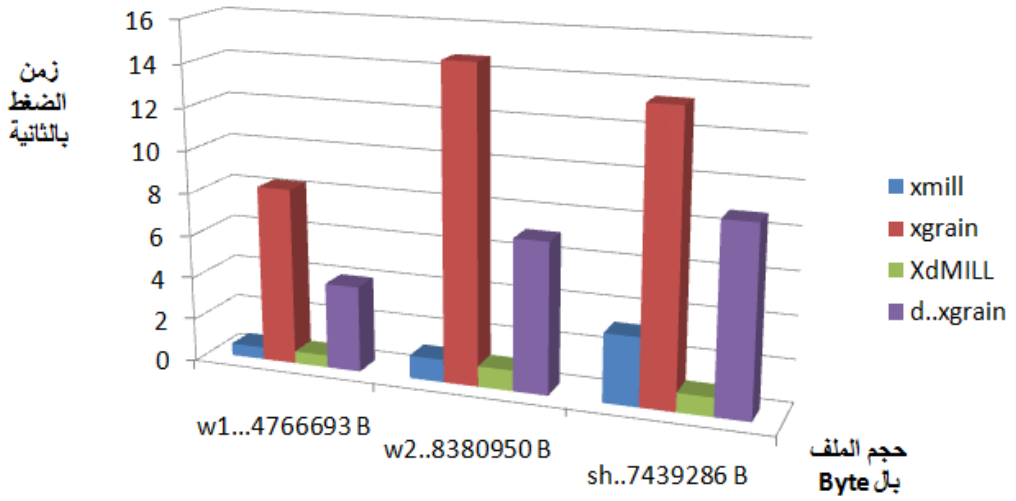
تم تطبيق كل من تقنيتي Xmill و XGrind على العينات السابقة وحساب زمن الضغط بالثانية وحساب زمن فك الضغط لكل تقنية الجدولين **(10)** و**(11)** و **الشكل (9)**.

الجدول (10) نتائج تطبيق طريقتي XGrind و Xmill بينها بالنسبة لزمن الضغط

الملف	حجم الملف الأصلي (in bytes)	زمن الضغطCT XMill	زمن الضغطCT XGrind
weblog1	4746593	0.57	8.30
weblog2	8387865	1.07	14.67
Shakespeare	9434567	3.21	13.41

الجدول (12) نتائج تطبيق طريقتي XGrind و Xmill بينها بالنسبة لزمن فك الضغط

الملف	حجم الملف الأصلي (in bytes)	زمن فك الضط XdMill	زمن فك الضط XGrind
weblog1	4746593	0.57	4.03
weblog2	8387865	0.97	7.06
shakespeare	9434567	0.89	8.75



الشكل (9): مقارنة زمن الضغط و فك الضغط لكل من XGrind و Xmill

- من النتائج الرقمية والبيانية يظهر أن زمن الضغط وفك الضغط في طريقة XGrind أكبر منه في مثيلتهاXMill.وبالتالي إن طريقة XMill أفضل من طريقةXGrind من حيث زمن الضغط للملفات المضغوطة.

الاستنتاجات والتوصيات:

الاستنتاجات:

- نلاحظ أن زمن الضغط باستخدام تقنية XGrind يساوي تقريباً ضعفي الزمن في تقنية Xmill
- طريقة الضغط GZip توفر معدل ضغط حوالي (50٪) حسب حجم الملف الأصلي.
- زمن إرسال الملفات المضغوطة بطريقة XMill أقل من زمن إرسال نفس الملفات مضغوطة بتقنيتي XGrind.
- طريقة الضغط Xmill تعطي نتيجة خوارزمية GZIP نفسها في حال حجم الملف المدخل المراد ضغطه أقل من 20KB وحسب النتائج يفضل استخدام طريقة XMillعلى ملفات حجمها أكبر من 1MB.

التوصيات:

العمل على استخدام أكثر من خوارزمية ضغط في حال أن حجم الملفات المراد إرسالها كبير.

المراجع:

1. SOHAIL ANSARI ; PRAJEET SHARMA , *XML Optimization and Compression* , International Journal of Innovations & Advancement in Computer Science , March 2015

2. SHERIF SAKR , *Investigate state-of-the-art XML compression techniques* , IBM Corporation , 19 July 2011

3. WILFRED NG ; WAI-YEUNG LAM ; JAMES CHENG , *Comparative Analysis of XML Compression Technologies* , March 2006.

4. HARTMUT LIEFKE ; DAN SUCIU , XMILL: *an Efficient Compressor for XML Data* , ACM New York, NY, USA , 2000.

5. ALEXANDER RIDGE WELL , *A Method for Efficient Transmission of XML Data across a Network* , University of Canberra (UC), Australia , 2007.

6. PANKAJ M. TOLANI & JAYANT R. *Haritsa / XGRIND: A Query-friendly XML Compressor* , IEEE Computer society , 2002

7. VOJTECH TOMAN , *Compression of XML Data* , Charles University, Master's Thesis at Department of Software Engineering Prague, 20 March 2003

8. SHERIF SAKR , *XML compression techniques: A survey and comparison* , Elsevier Inc , 2009 .

9. MARC GIRARDOT ; NEEL SUNDARES , AN MILLAU: *an encoding format for efficient representation and exchange of XML over the Web,* Elsevier B.V, 2000.

10. YEFEI XIN ; ZHEN HE ; JINLI CAO , *Effective Pruning for XML Structural Match Queries* , 2010.