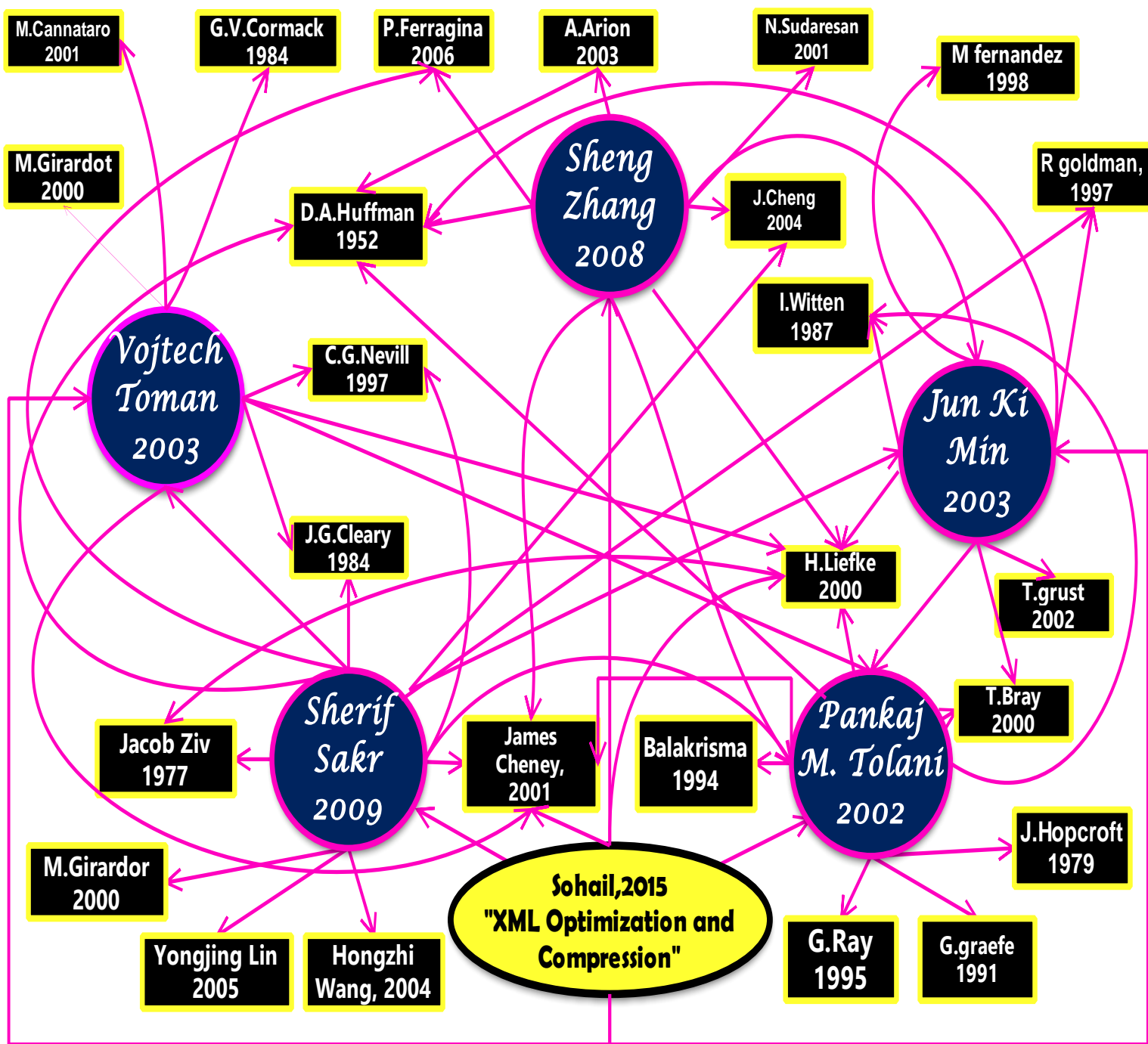


Tugas membuat Site map Citasi Paper Teknik Penulisan Karya Ilmiah



Donny Giovana Karo-Karo
09011181520011
SK 2a

Fakultas Ilmu Komputer
Universitas Sriwijaya



XML Optimization and Compression

XML(Extensible Markup Language) menjadi standar de-facto untuk bertukar dan menyajikan informasi tentang web. XML dengan cepat menjadi format standar de facto untuk data elektronik penataan, penyimpanan dan pertukaran. Telah banyak diadopsi di berbagai bidang industri komputer; misalnya, digunakan dalam berbagai database dan eBusiness aplikasi. Banyak harapan yang dibuat XML, dan oleh karena itu kemungkinan jumlah data yang tersedia di XML akan tumbuh secara substansial dalam waktu dekat. Namun, data XML juga diakui bertele tele, karena sangat memperluas ukuran data dengan tag dan struktur yang diulang. Masalah Data bertele-tele menimbulkan banyak tantangan pengolahan query dan pertukaran data. Ukuran ini membuat fakta bahwa jumlah informasi yang harus dipindahkan, diproses, disimpan, dan bertanya sering lebih besar dari format data lainnya. Ini bisa menjadi masalah serius di banyak kesempatan, karena data harus ditransmisikan dengan cepat dan disimpan teratur. Dokumen XML besar tidak hanya mengkonsumsi waktu transmisi, tetapi juga mengkonsumsi sejumlah besar ruang penyimpanan XML. Masalah ini dapat diatasi jika teknik kompresi XML digunakan untuk mengurangi kebutuhan ruang. Ada dua jenis kompresor XML. Teks Umum kompresor berdasarkan XML ringkas kompresor. Dari paper Sohail Ansari,2015 yang berjudul "XML Optimization and Compression" mencitasi paper dari Vojtech Toman,2003 " Compression of XML Data" menyimpulkan XML menjadi format standar untuk penyimpanan data elektronik dan pertukaran. Namun, kebutuhan ruang yang sering substansial lebih besar dibandingkan dengan format data lainnya. Masalahnya dapat diatasi jika data yang kompresi diterapkan. Kemungkinan kompresi sintaksis data XML berdasarkan pemodelan probabilistik struktur XML. Skema kompresi yang diusulkan bekerja secara berurutan, membuat kompresi dan dekompresi mungkin dilakukan. Selama masa dekompresi, data direkonstruksi dapat diakses melalui antarmuka SAX. Kami telah menerapkan perpustakaan kompresi XML eksperimental (Exalt) dan telah diuji secara luas kinerjanya pada berbagai dokumen XML. Paper Sohail Ansari,2015 "XML Optimization and Compression" mencitasi paper dari Sherif Sakr,2009"XML Compression techniques:A survey and comparison" menyimpulkan XML telah diakui sebagai standar de facto untuk representasi data dan pertukaran melalui World WideWeb(WWW). Menjadikan XML fleksibilitas yang besar dan penerimaan luas tapi di sisi lain itu adalah penyebab kelemahan utama yang menjadi besar dalam ukuran. Ukuran dokumen besar berarti jumlah informasi yang harus ditransmisikan, diproses, disimpan, dan bertanya sering lebih besar dari format data lainnya. Beberapa teknik kompresi XML telah diperkenalkan untuk menangani masalah ini. Dalam tulisan ini, kami menyediakan survei penuh atas teknik kompresi XML. Selain itu, kami menyajikan studi eksperimental yang luas dari implementasi yang tersedia dari teknik ini. Kami menyediakan sembilan kompresor XML menggunakan corpus besar dokumen XML yang meliputi sifat yang berbeda dan skala dokumen XML. Selain menilai dan membandingkan karakteristik kinerja dari alat kompresi XML dievaluasi, studi ini juga mencoba untuk menilai efektivitas dan kepraktisan menggunakan alat ini di dunia nyata.

Akhirnya, kami menyediakan beberapa panduan dan rekomendasi yang berguna untuk membantu pengembang dan pengguna untuk membuat keputusan yang efektif terhadap memilih alat kompresi XML yang paling cocok untuk kebutuhan mereka. Paper Sohail Ansari,2015 "XML Optimization and Compression" mencitasi paper dari Pankaj M. Tolani,2002 "XGRIND: A Query-friendly XML Compressor" yang menyimpulkan dokumen XML sangat bertele-tele sejak "skema" diulang untuk setiap record dalam dokumen. Sementara berbagai kompresor yang tersedia untuk mengatasi masalah ini, mereka tidak dirancang untuk mendukung pertanyaan langsung dari dokumen terkompresi, fitur yang berguna dari perspektif basis data. Dalam tulisan ini, kami mengusulkan sebuah alat kompresi baru yang disebut XGrind, yang secara langsung mendukung permintaan dalam domain terkompresi. Sebuah fitur khusus dari XGrind adalah bahwa dokumen terkompresi mempertahankan struktur dokumen, memungkinkan penggunaan kembali menggunakan teknik XML atau teknik standar untuk memproses dokumen terkompresi. Evaluasi kinerja yang lebih beragam dokumen XML dan permintaan pengguna menunjukkan bahwa XGrind secara bersamaan telah dijalankan ditingkatkan permintaan pengolahan dan perlakuan yang wajar. Paper Sohail Ansari,2015 "XML Optimization and Compression" mencitasi paper dari Jun- Ki Min,2003 "XPRESS: A Queriable Compression for XML Data" yang menyimpulkan HTML, banyak dokumen XML yang menetap di sistem file asli. Sejak XML data yang tidak teratur dan bertele tele, ruang disk dan bandwidth jaringan yang terbuang. Untuk mengakhiri ini, penelitian tentang kompresor untuk data XML telah dilakukan. Namun, beberapa XML presesor komputer tidak mendukung perhitungan data dikompresi, sedangkan kompresor XML lain yang mendukung perhitungan data dikompresi keseluruhan dengan tag dan nilai-nilai data menggunakan metode encoding yang telah ditetapkan. Dengan demikian, kinerja perhitungan data XML di komputer terdegradasi. Dalam tulisan ini, kami mengusulkan XPRESS, kompresor XML yang mendukung evaluasi langsung dan efisien perhitungan data XML terkompresi. XPRESS mengadopsi metode encoding, disebut kebalikan encoding aritmatika, cenderung untuk encoding jalur label data XML, dan menerapkan metode encoding beragam tergantung pada jenis nilai data. Hasil eksperimen dengan kehidupan nyata menunjukkan bahwa XPRESS mencapai perbaikan yang signifikan pada kinerja perhitungan untuk data XML dikompresi menjadi perhitungan yang wajar. Rata-rata, kinerja perhitungan XPRESS 2.83 kali lebih baik dari kompresor XML yang ada dan rasio kompresi XPRESS adalah 73%. Paper Sohail Ansari,2015 "XML Optimization and Compression" mencitasi paper dari Sheng Zhang,2008 " Comparative Research of XML Compression Technologies" menyimpulkan bahwa XML adalah standar de-facto untuk bertukar dan penyajian informasi di Web. Namun, data XML juga diakui bertele tele karena berat mengembangkan ukuran data yang menggunakan tag dan struktur diulang. Masalah Data bertele-tele menimbulkan banyak tantangan pengolahan permintaan dan pertukaran data konvensional. Halangan yang lebih jelas dalam perangkat bandwidth dan memori terbatas. Teknik kompresi adalah cara penting untuk mengatasi masalah bertele-tele. Makalah ini membandingkan dan analisis baru-baru ini mengusulkan teknologi kompresi XML, mempelajari pro dan kontra, kemudian mendiskusikan karya lanjutan kunci Kata XML,

kompresi data, pemrosesan perhitungan aplikasi Web. Sohail Ansari tidak hanya mencitai ke 5 paper diatas namun Ia mencitai beberapa paper yang dicitasi juga oleh lima paper diatas. Dari semua paper yang sudah dicitasi oleh Sohail Ansari menyimpulkan XML menjadi standar de-facto untuk bertukar dan menyajikan informasi tentang web. Namun, data XML juga diakui sebagai verbose, karena sangat memperluas ukuran data dengan tag dan struktur diulang. Masalah Data bertele-tele menimbulkan banyak tantangan pengolahan query dan pertukaran data. Halangan yang lebih jelas dalam perangkat bandwidth dan memori terbatas. Teknik kompresi adalah cara penting untuk mengatasi permasalahan yang bertele tele. XMill mengambil keuntungan dari kesadaran XML, sehingga membuat XMill sebagai alat yang efisien untuk mengompresi Data XML. Tiga pelaku yang digunakan dalam XMill untuk kompresi data, memisahkan struktur XML dari data, data grup terkait dan menerapkan kompresor semantik. The XMill menggunakan perpustakaan gzip untuk kompres data string XML untuk meningkatkan rasio kompresi. Metode baru yang diusulkan meningkatkan rasio kompresi alat XMill dengan menambahkan perpustakaan 7Zip di dalamnya. Dalam metode baru ini 7Zip perpustakaan menggunakan algoritma LZMA untuk kompres data. LZMA adalah versi yang disempurnakan & ditingkatkan dari algoritma LZ77 yang digunakan di perpustakaan gzip. Algoritma LZMA memiliki fitur berikut selama algoritma LZ77

- Menggunakan hingga 4GB panjang kamus bukannya 32KB untuk menghapus data ganda.
- Menggunakan pendekatan tampilan kepala bukan pendekatan serakah.
- Menggunakan parsing optimal, kode pendek baru-baru ini diulang di pengujian.
- Menggunakan penanganan konteks. Hasil penelitian dan analisis kinerja menunjukkan bahwa pendekatan yang diusulkan baru mencapai rasio kompresi yang terbaik dengan kecepatan kompresi sebanding. 5 Paper yang dicitasi oleh sohail,2015 juga mencitai masing masing 5 paper. Sheng zhang,2008 mencitai A.arion 2003 "XQueC: Pushing queries to compressed XML data" , N.sudaresan,2001" Algorithms and programming models for efficient representation of XML for internet applications", J.cheng,2004" XQzip: Querying compressed XML using structural indexing" , D.A Huffman,1952"A method for construction of minimum-redundancy codes",P. Ferragina, 2005"Structuring labeled trees for optimal succinctness, and beyond". Vojtech toman,2003 mencitai C. G. Nevill,1997"Compression and Explanation Using Hierarchical Grammars", J. G. Cleary,1984"Data Compression Using Adaptive Coding and Partial String Matching", M. Girardot,2000"An Encoding Format for Efficient Representation and Exchange of XML", M. Cannataro,2001"Compression of XML Data", G. V. Cormack,1984"Data Compression Using Dynamic Markov modelling". Sherif sark,2009 mencitai M.Girardor,2000" An encoding format for efficient representation and exchange of XML", Jacob Zib,1977"A universal algorithm for sequential data compression", Yongjin Lin,2005"Supporting efficient query processing on compressed XML",Hongzhin Wang,2004"Xcpaqs: Compression of XML document with Xpath query support",j.cheney,2001"Compressing XML with multiplexed hierarchical PPM models". Pankaj M tolani,2002 mencitai paper G. Ray,1995"Database Compression: A Performance Enhancement Tool", G.Graefe1991"Data Compression and Database Performance", J.

Hopcroft,1979"Introduction to Automata Theory, Languages, and Computation", T. Bray,2000"Extensible Markup Language (XML)", B. Iyerand,1994"Data Compression Support in Databases". Jun Ki Min,2003 mencitasi paper R. Goldman,1997"DataGuides: Enable Query Formulation and Optimization in Semistructured DataBases", M. F. Fernandez,1998"Optimizing Regular Path Expressions Using Graph Schemas", H. Liefke,2000"An Efficient Compressor for XML Data", I. H. Witten,1987"Arithmetic Coding for Data Compression", T. Grust,2002" Accelerating XPath Location Steps". Di dalam mind map yang saya buat ternyata Sohail tidak hanya mencitasi 5 paper namun ia juga mencitasi paper yang dicitasi oleh sherif sark yaitu paper james cheney,2001 "Compressing XML with Multiplexed Hierarchical PPM Models" dan paper yang dicitasi oleh Pankaj M tolani dan Jun ki min yaitu paper H.Liefke,2000 "XMill: An Efficient Compressor for XML Data". Pankaj M tolani juga mencitasi Paper yang dicitasi oleh sherif sark yaitu paper james cheney,2001 "Compressing XML with Multiplexed Hierarchical PPM Models" dan paper Sheng Zhang yaitu paper D.A Huffman,1952"A method for construction of minimum-redundancy codes". Jun Ki Min,2003 juga mencitasi Paper Pankaj M Tolani,2002 "XGRIND: A query-friendly XML compressor," dan juga paper yang dicitasi oleh Pankaj M Tolani yaitu paper T. Bray,2000"Extensible Markup Language (XML)", paper yang dicitasi oleh Sheng Zhang yaitu paper D.A Huffman,1952"A method for construction of minimum-redundancy codes". Sheng Zhang,2008 mencitasi Paper Jun Ki Min,2003 "XPRESS: A querable compression for XML data," Pankaj M Tolani,2002 "XGRIND: A query-friendly XML compressor," dan paper yang dicitasi oleh Sherif Sark yaitu paper James Cheney,2001"Compressing XML with Multiplexed Hierarchical PPM Models" dan paper yang dicitasi oleh M.tolani yaitu paper H.Liefke, 2000 "XMill: An Efficient Compressor for XML Data". Vojtech Toman,2003 mencitasi paper yang dicitasi oleh M.Tolani yaitu paper H.Liefke, 2000 "XMill: An Efficient Compressor for XML Data". Pankaj M Tolani,2002 "XGRIND: A query-friendly XML compressor," dan James Cheney,2001"Compressing XML with Multiplexed Hierarchical PPM Models". Sherif Sark,2009 mencitasi paper Vojtech Toman 2003 "Compression of XML Data," dan juga paper yang dicitasi Vojtech toman yaitu C.G neville,1997"Compression and Explanation Using Hierarchical Grammars" dan paper yang dicitasi oleh Sheng Zhang yaitu paper D.A Huffman,1952"A method for construction of minimum-redundancy codes" dan paper P. Ferragina, 2005"Structuring labeled trees for optimal succinctness, and beyond" dan J.cheng,2004" XQzip: Querying compressed XML using structural indexing". Ia juga mencitasi paper Jun Ki Min,2003 "XPRESS: A querable compression for XML data," dan juga paper yang dicitasi oleh Jun Ki mIn R. Goldman,1997"DataGuides: Enable Query Formulation and Optimization in Semistructured DataBases". Ia juga mencitasi Paper Pankaj M tolani,2002"XGRIND: A query-friendly XML compressor". Inovasi utama dalam mekanisme kompresi XML disajikan dalam pelaksanaan pertama di domain ini dengan Xmill. • kompresor Skema tergantung di praktek yang dominan di sebagian besar kompresor XML.