

組み込み機器で動作する 軽量のXMLデータベース

搭載メモリの少ないリソース環境下 でも高速検索を実現

XML (Extensible Markup Language) は、データの意味や構造を記述するためのマークアップ言語です。近年、カーナビや、モバイル端末、テレビ(TV)、AV機器などのデジタル機器においても、柔軟で拡張性の高いXMLデータの利用拡大が期待されています。しかし、従来のXMLデータベース(DB)は、搭載メモリの容量に制約があるデジタル機器では、十分な登録性能や検索性能が出ないという問題がありました。

東芝はこのような課題を解消するため、独自のXML符号化格納技術と、XQuery (XML Query Language)^(注1) コンパイル技術を開発し、XMLDBの軽量化に成功しました。

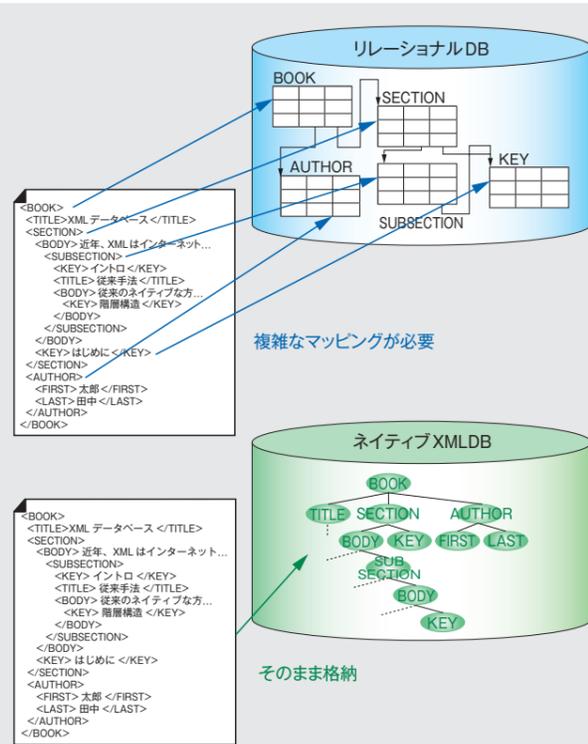


図1. ネイティブXMLDBのメリット—XMLデータの管理には、マッピングが必要なりレーショナルDBよりも、事前のデータ定義をせずに管理できるネイティブXMLDBが適しています。

デジタル機器における XML化の進展

データ構造の柔軟性や拡張性の高さから、近年、電子政府をはじめとして、電子商取引、事務文書管理、新聞コンテンツ管理などインターネット上の様々なデータ交換やデータ管理の領域で、XMLの利用が進んでいます。また、DLNA (Digital Living Network Alliance)^(注2) や、RFID (Radio Frequency Identification) PML (Physical Markup Language)^(注3)、RSS (Rich Site Summary)^(注4) など、ユビキタスやメタデータ(データについて)

(注1) XMLデータを検索するための言語。
(注2) 機器間の相互接続に関する技術仕様を策定する団体。
(注3) RFタグに埋め込まれた物体のID情報を記述するための言語。
(注4) Webコンテンツのメタデータを記述するための言語。

て関連情報を記述したデータの関連規格フォーマットにXMLが採用されています。更に、高性能・多機能化が進むデジタル機器においても、XMLデータの利用拡大が期待されています。

組み込み機器での動作を 実現する軽量のXMLDB

柔軟性や拡張性の高いXMLデータの管理には、表形式のデータの処理に適したリレーショナルDBよりも、事前のデータ定義をせずに多様な構造データを管理できるネイティブXMLDBが適しています(図1)。しかし、搭載メモリの容量に制約があるデジタル機器では、リレーショナルDBと比較して十分な登録性能や検索性能が出ないという問題がありました。

この問題を解決するために、XMLの構造データをコンパクトな数値列で表

現し、その数値列に対する1回の走査だけで構造データに対する検索処理を行う、軽量のXMLDBを開発しました。そのコア技術となるのが、XML符号化格納技術とXQueryコンパイル技術です(図2)。

ツリー構造の情報を大幅に 削減するXML符号化格納技術

一般的なXMLDBでは、構造データをそのままツリーの形でDBに格納するため、ツリーを表現するデータ量が元のXMLデータの数倍にも膨れ上がるという問題がありました。XML符号化格納技術では、XMLの各構造要素に対して1個の数値を割り当てて1次元の数値列を生成し、更に、その数値列に対して圧縮をかけた符号列をDBに格納します。構造要素を1個の数値で表現していることと、規則性の

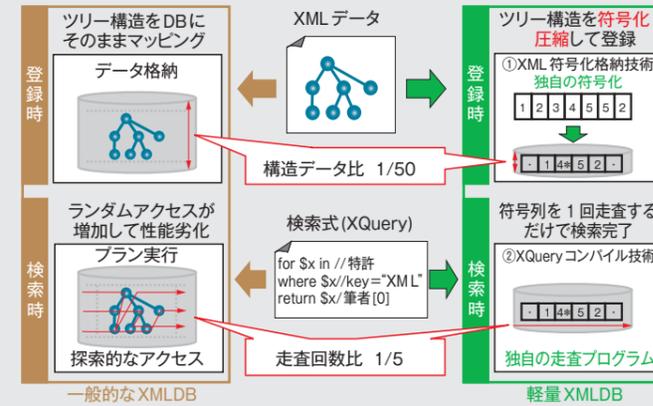
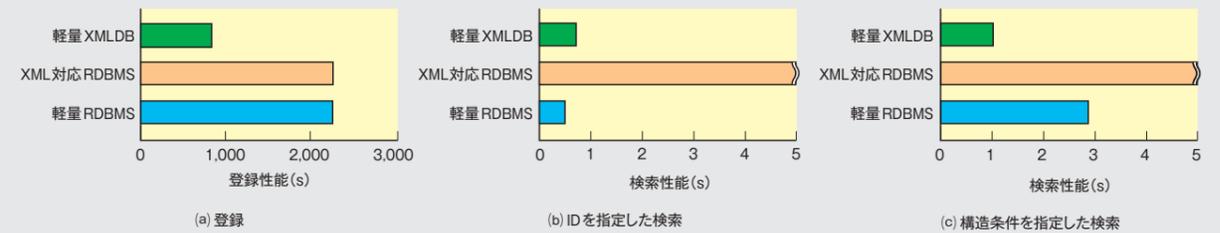


図2. 軽量XMLDBのコア技術—軽量性と高速性を達成するため、XML符号化格納技術とXQueryコンパイル技術を開発しました。



RDBMS: リレーショナルDB管理システム

図3. 他データベースとのベンチマーク—標準ベンチマークデータセットを使って、1Gバイトのデータに対する登録と検索の処理時間を計測しました。

ある数値列は通常大幅に圧縮できるという2点の効果で、XMLデータを文書オブジェクトモデル(DOM)で表現する場合に比べて、構造データのサイズを約1/50にまで削減できました。

検索を高速に実行する XQueryコンパイル技術

XMLDBへの検索には通常XQueryと呼ばれる問合せ言語を使います。XQueryでは構造に対する条件を記述できますが、構造条件が複雑になるとXMLデータの構造要素に対するアクセスが増大します。その結果、少リソース環境下ではデータの多くがメモリに乗り切れないため、ハードディスクなどへのアクセスが大量に発生して、検索速度が大幅に低下します。

XQueryコンパイル技術は、XQueryで表現された検索命令を効率の良い内

部コード(走査プログラム)に変換する技術です。内部コードは、先に述べた符号列への走査手順を示しており、XQueryに出現する構造条件の数によらず、常に1回のアクセスで検索を終了します。

これらの技術により、メモリ使用量がわずか数Mバイトの条件下でも、一般的なXMLDBと比べて検索性能を10倍以上向上させることができました。また、図3に示すように、リレーショナルDBと比較しても同等以上の性能が得られています。

実際にモバイル端末上でXMLDBを動作させているようすを図4に示します。

将来への展望

XMLDBが持つ高い柔軟性を維持しつつ、軽量性及び高速性を達成しました。軽量のXMLDBをデジタル機器に



図4. モバイル端末上での動作—数Mバイトのメモリ使用量で、数百MバイトのXMLデータを検索することができます。

組み込むことで、XMLのメリットを生かした多様なコンテンツの検索や活用が可能になります。例えば、このXMLDBをデジタルTVに組み込めば、番組コンテンツに関する番組情報やシーン情報などを蓄積しておき、出演者やシーンに関連した様々なキーワードで見たいシーンを指定して、それらの一覧を瞬時に表示するといった機能が実現できると考えています。

今後は、DBとしての必要な機能を充実させ、早期の実用化を目指します。

服部 雅一

研究開発センター
知識メディアラボラトリー研究主幹