

電子商取引における XML 検索・変換の最適化方式

春日史朗 鳥海幸輝 坂田哲夫 小林伸幸 芳西崇

E-mail: {kasuga, koki, sakata, kobayasi, honishi}@dq.isl.ntt.co.jp

日本電信電話株式会社 NTT サイバースペース研究所

あらまし

電子商取引において、企業間で XML 電子伝票をやり取りするためには、共通スキーマを用いる必要がある。しかし、各企業の DB に格納された XML 電子伝票は、各企業の独自スキーマで記述されるため、多量の XML 電子伝票を独自スキーマと共通スキーマ間で構造変換する必要がある。従来、XML の構造変換は DB の外部で行われており、DB 検索と、検索結果の構造変換は逐次的に処理されていた。そのため、DB 検索の結果に、構造変換後に不要な情報が含まれることになり、DB 検索と構造変換のトータルコストが上昇するという問題があった。本稿ではこの問題に対し、共通スキーマを XML ビューとして定義し、XML ビューと DB のスキーマとの間の構造変換の際に、構造変換に最適化した DB 検索を行うことで、DB 検索と構造変換のトータルコストを低減する方式を提案し、従来方式との比較を通して、本手法の有効性を実証する。

キーワード

情報検索, データベース, XML, 電子商取引

An Optimization of Retrieving and Translating XML Data for Electronic Commerce

Shiro KASUGA, Koki TORIUMI, Tetsuo SAKATA, Nobuyuki KOBAYASHI
and Takashi HONISHI

E-mail: {kasuga, koki, sakata, kobayasi, honishi}@dq.isl.ntt.co.jp

NTT Cyber Space Laboratories, NTT Corporation

Abstract

In B2B Electronic Commerce, it is necessary to use a common schema to exchange XML documents between the enterprises. However, XML documents are described in an original schema of each enterprise. So it is needed to convert XML documents from each enterprise's schema to common one. In existing systems, XML document conversion is done outside of a B2B framework. Therefore retrieving and translating XML documents are processed sequentially, so it increases total cost of the XML data processing.

In this paper, we propose the optimization of retrieving and translating XML documents for B2B Electronic Commerce, using XML view.

Keyword

information retrieval, database, XML, Electronic Commerce

1. はじめに

近年、情報流通のフォーマットとしてXML (eXtensible Markup Language)^[1]が注目を集めている。XMLは優れたデータ表現力と拡張性を備えており、企業間の電子商取引(EC)などへの応用が期待されている。

電子商取引において、企業間でXML電子伝票をやり取りするためには、共通スキーマ^{[2][3]}を用いる必要がある。しかし、各企業のデータベースに格納されたXML電子伝票は、各企業の独自スキーマで記述されることが多いため、多量のXML電子伝票を独自スキーマと共通スキーマ間で構造変換する必要がある。従来、XMLの構造変換はデータベースの外部で行われており、データベース検索と、検索結果の構造変換は逐次的に処理されていた。そのため、データベース検索の結果に、構造変換後には不要な情報が含まれることになり、データベース検索と構造変換のトータルなコストが上昇するという問題があった。

本稿ではこの問題に対し、各企業の独自スキーマを共通スキーマに対するXMLビューとして定義し、XMLビューとデータベースのスキーマ(以下、蓄積スキーマ)との間の構造変換の際に、構造変換に最適化したデータベース検索を行うことで、データベース検索と構造変換のトータルなコストを低減する方式を提案する。

2. 既存方式と問題点

現在、中小規模のB2B電子商取引や、マーケットプレースのプラットフォームとして、図1に示す、集中型のB2Bプラットフォームが用いられている。

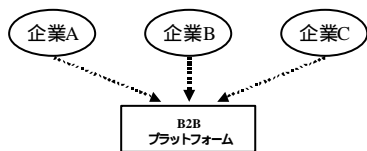


図1 集中型 B2B 電子商取引の構成

ここで、各企業のデータベースに格納されたXML電子伝票は、各企業の独自スキーマで記述されるため、XML電子伝票のスキーマを、各企業の独自スキ

ーマから、共通スキーマに変換する必要がある。現在、XML電子伝票をスキーマ変換する方式として、各企業側でXML変換エンジンを用いて変換し、B2Bプラットフォームと送受信する方式が広く用いられている(図2)。一般的にXML変換は、W3Cにより標準化された、XML構造変換言語XSLT^[4]により記述でき、現在、多くのXSLTによるXML変換処理系が実装されている。

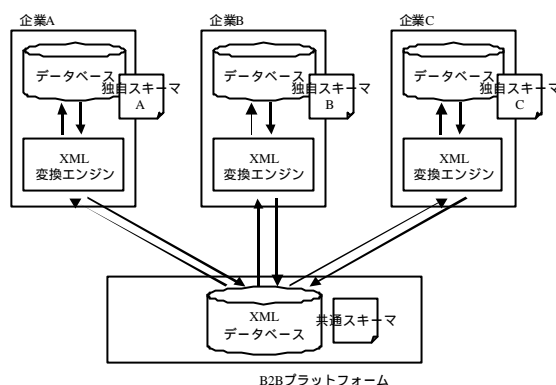


図2 既存の B2B プラットフォーム

既存方式は、XMLデータベースとXSLTによる変換エンジンが独立しているため、各企業のユーザがB2Bプラットフォームにあるデータベースから、独自スキーマで検索結果を得ようとする場合、以下のような2段階の処理が必要であった。

手順 B2Bプラットフォームのデータベースに蓄積されているXML電子伝票のスキーマに合致した検索を行う

手順 その検索結果をXML変換エンジンにより変換して、最終的な検索結果を得る

しかし、この方式では、

問題点1 手順にてB2Bプラットフォームのスキーマで検索しており、プラットフォームのスキーマ変更が各企業のシステムに及ぶため、各企業のシステムと、B2Bプラットフォームとの独立性が悪い。

問題点2 共通スキーマは、各企業が必要とする情報を漏れなく盛り込むため、個々の企業にとっては、変換によって不要となるXMLの要素も

B2B プラットフォームからの検索結果として各企業へ出力されるため、余分な通信コストがかかる(図3)。

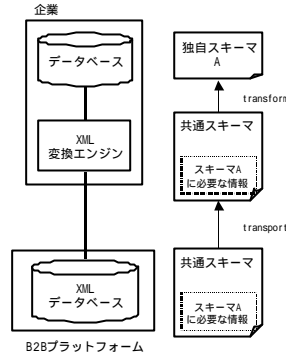


図3 既存方式における通信と変換

という問題があった。また、このような構造上の異種性に起因する問題に加えて、B2B電子商取引に用いられるXML電子伝票では、値の表現形式が多種多様であり、値の表現形式の変換機能が重要である。そのため、

問題点3 XML文書の構造変換記述言語として、XSLTが用いられることが多いが、記述が難しく、値の表現形式の変換機能が不足しており、電子商取引に対応できない。

という問題があった。

3. 提案する方式

前述の、既存方式における構造上の異種性、値の異種性に起因する問題点に対し、我々はXMLビューに基づくB2B電子商取引システム構築手法を提案する。

3.1. XML文書検索・変換システム

我々は、XMLデータベースに対して、XMLビューを定義し、XMLビューを通して、XML文書の検索、更新を行うことができるXML文書検索・変換システムを開発している。

本システムのXML文書変換機能は、蓄積スキーマとビューの間で、以下のようなスキーマ変換を可能とする。

● 構造の変換

図4に示すようなXML文書の木構造が変化する変換である。

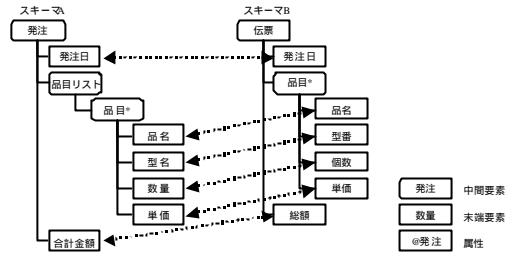


図4 構造の変換

● 値の変換

図5に示すような値の表現形式が変化する変換である。



図5 値の変換

本システムでは、以下の2つの方式により、この変換を実現する。

- 変換ルールによる構造変換方式
- ドメインによる値の変換方式

3.2. システム構成

XML文書検索・変換システムの構成を図6に示す。

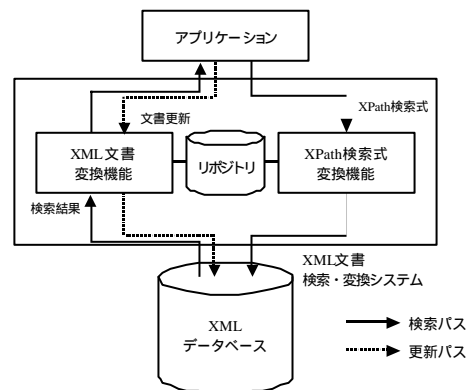


図6 XML文書検索・変換システム

本システムは、以下の要素から構成されている。

- リポジトリ
ビューの定義情報と変換情報とを保持する辞

書。

- XPath 検索式変換機能
アプリケーションから XML データベースに対して XML 文書検索を行う際に、XPath^[5]で記述された検索式の変換を行う機能である。本機能は、リポジトリの定義情報に従って変換処理を行う。
- XML 文書変換機能
アプリケーションから XML データベースへの XML 文書挿入・更新時、XML データベースからアプリケーションへの XML 文書検索結果返却時に、XML 文書の構造変換と値の変換を行う機能である。本機能は、リポジトリの定義情報に従って変換処理を行う。

これらの機能により、本システムは、蓄積スキーマで XML データベースに蓄積された XML 文書を、その利用目的やアプリケーション（以下、AP）に応じて、仮想的なスキーマであるビューに変換したように見せ、ビューに対して XML 文書の検索・更新を行うことができる。また、ビューを階層的に定義し、ビューに対するビューへの構造化文書の検索・更新を行うことが可能である。（図 7）

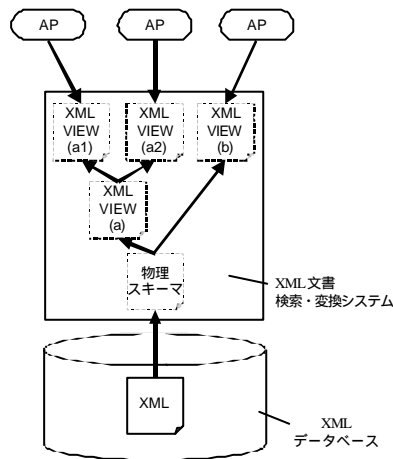


図 7 蓄積スキーマとビュー

3.3. 変換ルールによる構造変換方式

本方式では、XML 文書の構造変換において、要素間の対応関係（以下、変換ルール）を予めリポジトリに登録し、個々の XML 文書を変換ルールを用いて変換する^[6]。

変換ルールの例を図 8 に示す。図中、末端要素は値を持つ要素であり、中間要素は子要素を持つ要素である。値と子要素の両方を持つ要素は、一般的に、XML 電子伝票では使われないため、本システムでは未対応である。

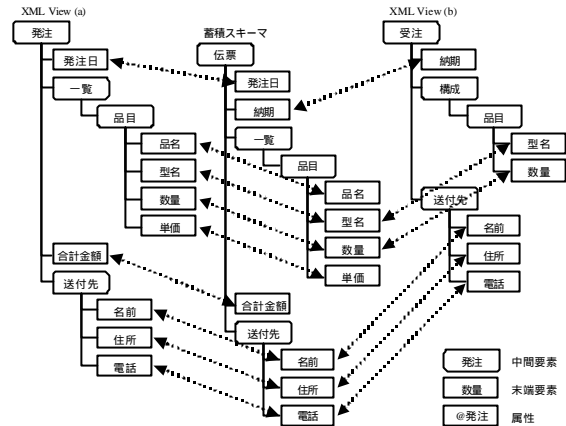


図 8 変換ルールの例

本システムでは、ビューを通して、XML 文書の検索、更新を可能とする。以下で、それぞれの処理手順を述べる（詳細な処理手順については付録として記載する）

XML 文書の更新

図 8 において、XML View(a)を通して、XML 文書を更新する場合、以下の手順により処理される

- Step1** AP により View(a)形式の XML 文書が入力される
- Step2** 変換ルールに従い、この XML 文書が、蓄積スキーマ形式の XML 文書に変換される。
- Step3** 変換された XML 文書により、XML データベースの既存文書が置き換えられる。（文書単位の置き換え、要素単位の置き換えが可能である）

なお、ビューの定義が多段の場合は、Step2 の処理が繰り返し行われる。

XML 文書の検索

図 8 において、XML View(b)を通して、XML 文書を検索する場合、以下の手順により処理される。

- Step1** AP により View(b)形式での XPath 検索式が入力される（例として `"/ 受注 [納期 = '2002/1/11']"`）
- Step2** 検索式が変換ルールに従って蓄積スキーマ

マ形式に変換される（"/伝票[納期='2002/1/11']"）

Step3 変換された検索式が XML データベースに適用され，検索結果を得る

Step4 検索結果は変換ルールにより，XML View(b)にスキーマ変換される

Step5 変換された検索結果がユーザに返却される

なお，ビューの定義が多段の場合は，Step2～ Step4の処理が繰り返し行われる。

本方式による検索処理では，XML View(b)を構成する際に不要な要素は，APに返却されないため，システムから APへの検索結果の通信に無駄がないことに注意されたい。

3.4. ドメインによる値の変換方式

値の変換を実現するため，本システムでは，表現形式の異種性を解消する方式^[7]を用いる。

日付の表記や重さの単位などは，表現の形式にいくつかの種類がある。そのため，複数のスキーマでは違った表現形式である場合が多く，XML 検索・変換を行う場合に障害となる。表現形式の変換には，表現形式の変換関数をすべて記述する手法が考えられるが，表現形式が n 種類ある場合，全ての表現形式変換を網羅すると，変換関数が n^2 個必要になるという問題がある。そこで本システムでは，ドメインという概念を用い， n 個の変換関数で全ての表現形式間の変換を実現する以下の方式を用いる。

ドメインの概念を図示すると図 9 のように表現できる。ここで，「¥1,000」や「千円」などのデータの 1 つの表現形式をそれぞれ 1 つのドメインとみなし，相互に変換可能なドメインの一群をドメイングループと呼ぶ。ドメイングループ中の 1 つドメイン（グローバルドメイン）を代表とし，他のドメインとの間の変換関数（ドメイン変換関数）を定義することで，全ての表現形式の相互変換を提供する。ドメインの定義は，システム内のリポジトリにドメイン名，ドメイングループ名，ドメイン間の変換関数を設定し，スキーマの要素毎にドメイン名を対応付けることで行う。この処理は前記の構造変換処理に併せて行う。

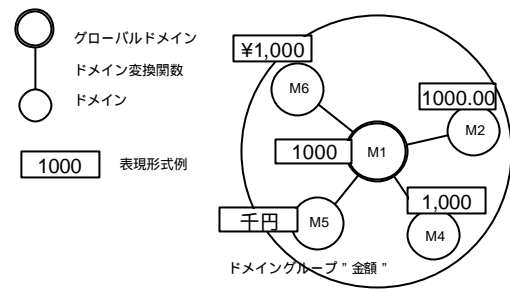


図 9 ドメイン

3.5. ビュー定義情報

前記の，変換ルールと，ドメインの情報は，管理者がリポジトリに設定する必要がある。ここで，変換ルールを作成することは，大きな稼働を必要とする作業である。しかし我々は，この問題に対して，変換ルールを要素名などから推定し，ユーザによる変換ルールの作成稼働を軽減する機能を提案している^[8]。

4. B2B 電子商取引システム

前述の XML 文書検索・変換システムを用いた，B2B 電子商取引システムの構成を図 10 に示す。

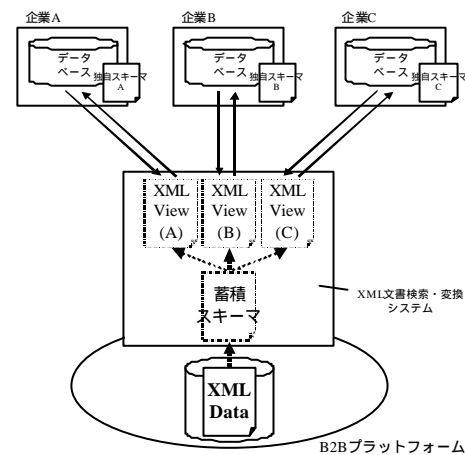


図 10 B2B 電子商取引システム

各企業は，B2B プラットフォームの XML データベースに対して，各企業の独自のビューを通してアクセスすることが可能である。これにより，各企業のデータベースの独自スキーマを独自のビューとして定義することで，容易に各企業のシステムと B2B プラットフォームを連携させることができ，B2B プ

ラットフォームと各企業間の通信は、各企業が必要とする量に抑えられ、通信コストの低減が可能となる。(図 11)

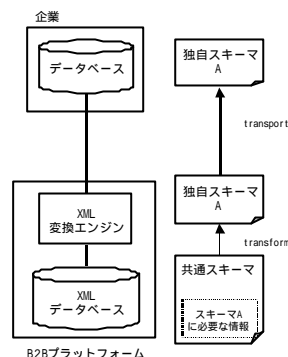


図 11 提案方式における通信と変換

5. 考察

我々は、XML 文書検索・変換システムを集中サーバに設置することで、B2B 電子商取引に必要な下記の機能が実現され、既存システムの問題点を解決することができることを示した。

- 各企業は独自スキーマをビューとして登録し、ビューを通して検索・更新を行うことができ、各企業の社内システムの独立性が向上する (問題点 1)
- 検索結果の出力に必要な部分だけをデータセンターから各企業に通信すればよいため、検索・変換処理のトータルコストが削減される (問題点 2)
- B2B 電子商取引に必要な不可欠な値の変換が実現される (問題点 3)

6. まとめと今後の課題

我々の開発している XML 文書検索・変換システムを適用することで、既存の B2B 電子商取引システムの問題点を解決する方式を提案した。

今後の課題としては、

- 変換ルールの作成コスト低減手法の検討
- 既存システムとのコスト比較の実施がある。

参考文献

[1] W3C, “Extensible Markup Language

(XML) 1.0,” <http://www.w3.org/TR/REC-XML>, 2000.

[2] RosettaNet, <http://www.rosettanet.org>.

[3] ebXML, <http://www.ebxml.org>.

[4] W3C, “XSL Transformations (XSLT) Version 1.0,” <http://www.w3.org/TR/xslt>, 1999.

[5] W3C, “XML Path Language (XPath) Version 1.0,” <http://www.w3.org/TR/xpath>, 1999.

[6] M. Onizuka and T. Honishi, “XTL: An XML transformation language,” Proc of Int. Conf. Extreme Markup Language, pp.167-180, 2001.

[7] 星野, 綱川, 町原: “DBSENA: マルチデータベース環境における情報資源管理と検索方式,” 第 144 回データベースシステム研究会, 1998.

[8] 鳥海, 春日, 坂田, 小林, 芳西, “情報流通分野における XML 変換方式の研究,” DEWS2002 (予定), 2002.

[9] 石川 博: “XML とデータベース, 情報処理,” Vol. 41, No. 1, pp. 68-73, 2000.

[10] 吉川正俊, 志村壮是, 植村俊亮: “オブジェクト関係データベースを用いた XML 文書の格納と検索,” 情報処理学会論文誌データベース, Vol. 40, No. SIG6(TOD3), pp. 115-131, 1999.

[11] S. Abiteboul, S. Cluet, V. Christophides, T. Milo, G. Moerkotte, and J. Siméon: “Querying Documents in Object Databases,” International Journal of Digital Libraries, Vol.1, No.1, pp.5-19, 1997.

付録 変換ルールによる構造変換方式

I XML 構造変換方式

XML 文書の構造変換について、XML 文書 D (図 12) から XML 文書 D' (図 13) への変換の例を用いて説明する。この時、ビュー中の要素と蓄積構造のスキーマ中の要素の対応関係である「変換ルール X (図 14)」が用意されているものとする。具体的な手順は以下のステップに従う。

```
<?xml version="1.0"?>
<商品>
  <品名>BIO2001</品名>
  <構成品>
    <種別>CPU</種別>
    <機能>800MHz</機能>
  </構成品>
  <構成品>
    <種別>メモリ</種別>
    <機能>256MB</機能>
  </構成品>
  <構成品>
    <種別>ハードディスク</種別>
    <機能>10GB</機能>
  </構成品>
  <構成品>
    <種別>ハードディスク</種別>
    <機能>20GB</機能>
  </構成品>
  <価格>¥150,000</価格>
</商品>
```

図 12 XML 文書 D の例

```
<?xml version="1.0"?>
<product>
  <name>BIO2001</name>
  <parts>
    <item type="CPU">
      800MHz
    </item>
    <item type="メモリ">
      256MB
    </item>
    <item type="ハードディスク">
      10GB
    </item>
    <item type="ハードディスク">
      20GB
    </item>
  </parts>
  <price>150,000</price>
</product>
```

図 13 変換された XML 文書 D' の例

ビュー	蓄積構造
/	/
/商品	/product
/商品/品名	/product/name
/商品/構成品/機能	/product/parts/item
/商品/構成品/種別	/product/parts/item/@type
/商品/価格	/product/price

図 14 変換ルール X の例

Step1 XML 文書の入力

変換元の XML 文書を入力する。この時、その XML 文書の木構造を図 15 と表すとする。ここで、木構造中の番号は、個々の要素に行きかけ順で付与した番号である。

Step2 構造変換

ビューに出現する要素に注目し、以下の手順を繰り返す。

Step2a パスの取り出し

注目する要素から、ルート要素に至るパスを、取得する。例えば、番号 1 の要素のパスは、「/商品/構成/種別」である。

Step2b パスの変換

前ステップで作成したパスをキーとして、「変換ルール X」より、変換先のパスを作成する。変換元のパス「/商品/構成/種別」をキーとすると、変換先のパス「/product/parts/item/@type」が得られる。なお、ここで、指定のキーが「変換ルール X」に無い場合は、このパスを無視して次の要素に進む。

Step3 変換されたパスの配置

変換先のパスを変換先の木構造中に配置する。例として、変換先のパス「/product/parts/item/@type」は、図 16 の 1 番の要素として配置する。このとき、パスの途中の要素が無い場合は、新たに作成する。図 16 の「/product/parts」要素は、要素 1 を配置する際に、同時に作成されたものである。

Step4 結果出力

Step2 より、図 17 の木構造で示される XML 文書が作成され、これを出力して本手順を終了する。

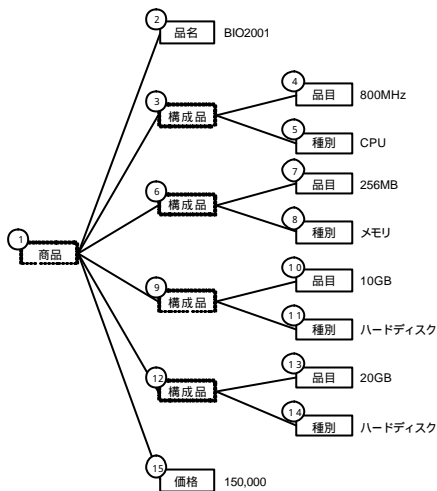


図 15 XML 文書 D の木構造

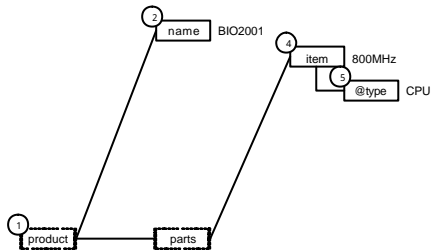


図 16 構造変換途中の XML 文書 D' の木構造

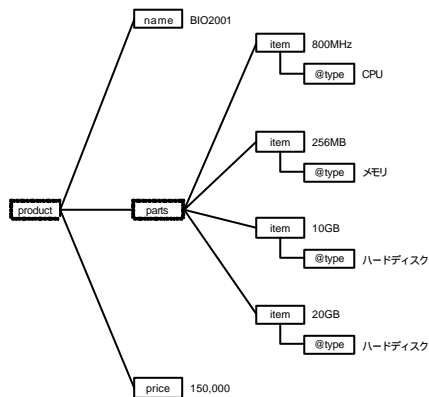


図 17 構造変換された XML 文書 D' の木構造

II. XPath 検索式変換方式

検索式の構造変換は、検索式を木構造ととらえ、ビューのスキーマを変換元のスキーマ、蓄積構造のスキーマを変換先のスキーマとし、XML 構造変換方式と同等の処理を行うことで実現できる。例として、

図 18 に示す「検索式 Q」は、図 19 に示す木構造ととらえ、図 20 に変換することで、図 21 に示す「構造変換された検索式 Q'」を得る。

/商品/構成部品[種別="CPU"]

図 18 検索式 Q

商品 構成部品 種別 = "CPU"

図 19 検索式 Q の木構造

product parts item @type = "CPU"

図 20 構造変換された検索式 Q' の木構造

/product/parts/item[@type="CPU"]

図 21 構造変換された検索式 Q'