データベースシステム 120－10
（２０００．１．２４）

入力要素を保存する集約に基づくビューへの問合せ最適化手法

加藤弘之† 吉川正俊§ 大山敬三† 植村俊亮§

†学術情報センター
§奈良先端科学技術大学院大学 情報科学研究科

内部に構造を有するデータを管理するデータベースにおいて，入力要素を保存する集約は自然に定義される．本稿では，XMLデータを対象として入力要素を保持する集約によって定義されるデータベースビューに対する問合せ処理の最適化手法を提案する．本稿で提案する最適化手法は非再帰問合せを対象とするマジックに基づく手法である．

A Query Optimization for Views Constructed by Aggregations Preserving Input Sets

Hiroyuki KATO†, Masatoshi YOSHIKAWA§, Keizo OYAMA† and Shunsuke UEMURA§

†National Center for Science Information Systems(NACSIS)
§Graduate School of Information Science, Nara Institute of Science and Technology (NAIST)

We propose a optimization method for queries to database views constructed by aggregations preserving input sets. This method is characterized by applying *Magicsetrewriting*, well known as a optimization based on "Sideways Information Passing" for non-recursive queries as well as recursive queries.

1. はじめに

内部に構造を有するデータを管理するデータベースにおいて，入力要素を保存する集約は自然に定義される．後で述べるように，例えば，XML 文書を管理するデータベースにおいて，そのような集約は有用である．

XML[WWWC98] は W3C(World Wide Web コンソーシアム) によって勧告された，内部に構造を有するデータを記述するためのメタ言語である．XML の重要な応用の一つが Web 上のデータの共通交換フォーマットである．異種分散環境における情報源のデータを XML でラップすることで，格納手法に依らない情報の統合が可能となる．

このような異種分散環境における情報源のデータの統合はデータベースビューとして捉えることができる．全ての情報源が問合せ最適化機構を支援しているわけではないので，XML データに対する論理的な問合せ最適化手法を開発することは有益なことである．

本稿では，XML データに対する論理的な問合せ最適化手法を提案する．この最適化手法は，伝統的なデータベース問合せの最適化手法であるマジック [AHV95] に基づくものである．本来マジックは，再帰問合せの bottom-up アプローチによる最適化手法として開発されたが，後で非再帰問合せである関係データベース問合せの最適化手法へと拡張された．本稿ではこの非再帰問合せに適用されたマジックを用いて XML データを操作する問合せの最適化を提案する．

以下，本稿は次のような構成である．この節の残りでは，本稿で提案する最適化手法を直観的に説明する．2 節では本稿における技術的な準備として，XML と XML-QL そして，マジックについてその概略を説明する．3 節はマジックを用いた XML データに対する問合せ最適化手法について記述する．4 節は本稿のまとめと今後の課題である．

1.1 研究成果の直観的説明

この節では，本稿で提案する最適化手法について直観的に説明する．XML フォーマットで新聞記事を配信しているサーバについて考える．このサーバは記事の管理には関係データベースと文書レポジトリを用いているかもしれないが，記事を配信する際は，XML でラップして配信することでクライアント側ではサーバが用いている格納管理機構に関わらず統一的な操作が可能となる．

サーバが提供している記事の見出しを日付でグループ化して連結したようなビュー `view1` について考える．このビューはクライアント側に配信されるが，クライアント側のなんらかの理由，例えば格納容量または過去の記事を保持する必要がないという理由で，保持されていないものとする．このような状況の元で，このビューに対する「日付でグループ化した見出し集合の中に"Clinton"が出現するものを検索せよ」という問合せ処理について考える．この問合せは，伝統的なデータベース問合せ処理手法においては，ビューを展開することで，以下のような手順で処理される．

i) サーバ側でビュー `view1` の定義を評価．

ii) 評価されたビュー `view1` 上で，"Clinton" を含むという条件を評価．

しかしながら，この手順ではまず無条件に `view1` のビュー定義を評価するので，"Clinton" を含まない見出しの日付に関しても，計算をしてしまう．これに対して，本稿で提案する最適化手法を適用すると次のような処理手順となる．

i) 見出しに"Clinton"が出現する記事の日付を特定．

ii) その日付に関する記事だけを用いて `view1` を評価．

この最適化手法は，後で記述するように本質的にはマジックに基づく手法である．これにより，問合せ結果に関連のないデータを用いた計算をする必要がないので，問合せ処理中の中間結果のデータ量を減すことができ，結果として問合せ処理全体の最適化につながることがある．もちろん，最終的には問合せ書き換えのためのコストも考慮に入れて判断する必要がある．

2. 準備

2.1 XML

XML は，W3C(World Wide Web Consortium) のもとで開発された構造化データを記述

サーバが提供する XML 文書の DTD

```
<!ELEMENT article
     (date, headline, body)>
```

view1 の DTD

```
<!ELEMENT headlines-by-date
     (date, headline+)>
```

図 2 DTD の例

するための言語である．もともと，文書を対象として開発された経緯により，XML で記述されたオブジェクトは，XML 文書と呼ばれるが，文書だけではなく，内部に論理構造を有するデータ全てがXMLの記述対象である．XMLの目標は，現在のHTMLと同様に，Web上で配布，受信，処理できるようにすることである[?]．XML 文書はエレメントと呼ばれるデータの論理構造を，開始タグと終了タグを明示的に文書中に埋め込むことで表現したデータオブジェクトである．

DTD は文書中のエレメントの出現の順序を規定したものであり，本質的な役割は文書の論理構造に関する文法とみなすことができる．図1中のサーバが提供している XML フォーマットの新聞記事の DTD を図 2に示す．

2.2 XML-QL

XML-QL[DFF+98, DFF+] は，XML 文書を操作するための問合せ言語であり，WHERE-CONSTRUCT 構成子を有している．XML-QL の WHERE 構成子は SQL の WHERE 構成子に，XML-QL の CONSTRUCT 構成子は SQL の SELECT 構成子に，それぞれ相当するものである．入れ子問合せは CONSTRUCT 句の入れ子によって表現される．図 3に，図 1に示したビュー `view1` に対する問合せを XML-QL で表現したものを示す．

2.2.1 利用者定義関数と利用者定義述語

本稿では，XML-QL に対して利用者定義集約関数 `cat()` と利用者定義述語 `CONTAIN` を導入する．`cat()` は入力集合の要素を連結するこ

ビュー定義のための問合せ

"www.localserver/headlines-by-date.xml"定義のための問合せ

```
WHERE <article>
        <date>$d</>
        <headline>$h</> ELEMENT_AS $e
      </>
      IN "www.newsserver/news.xml"
GROUP-BY $d
CONSTRUCT <headlines-by-date>
            <date>$d</date>
            <headlines>
             cat($e)
            </></>
```

Main Query Block

```
WHERE <headlines-by-date>$hs</>
                ELEMENT_AS $ans
      IN "www.localserver/h-by-d.xml",
      $hs CONTAIN 'Clinton'
CONSTRUCT $ans
```

図 3 XML-QL 問合せの例

とで集約する．図 4は従来の集約と本稿で導入する `cat()` との違いを示している．この図から明らかなように従来の集約が入力要素を保持しないのに対して，`cat()` は入力要素を保持する集約である．また，述語 `CONTAIN` は 2 項演算子であり，与えられた文書中に与えられた文字列が出現するときに限り，真を返すものとする．

2.3 マジック

マジックは，演繹データベースの分野において，論理プログラミングにおける手法を採り入れた再帰問合せのための最適化手法として開発された [BMSU86, BR87]．その手法は，SLD-resolution(Selection rule-driven Linear resolution for Definite clauses) を set-at-a-time に適用したものであり，bottom-up アプローチをとっている．本質的には，SIPS(Sideways Information Passing Strategy) により，bf adorn-

ments[1]のついたsupplementary関係を用いて，問合せ処理中において，答えに直接結びつかない余分な組の生成を抑えることで問合せ最適化を行なうものである[AHV95].

その後，非再帰問合せに適用され，関係データベース問合せの最適化手法としても確立されたものとなった[MFPR90]．本稿では，この非再帰問合せに適用されたマジックを用いた最適化手法について記述する．

3. XML-QLへのマジックの適用

この節では，XML-QL問合せに対してマジックがどのように適用されるかについて例を用いて説明する．既に述べたように，図1に示したXMLビューがクライアントサイトに配信後に何らかの理由で保存されていない場合について考える．このとき，伝統的なデータベース問合せ処理を適用すると，図3に示したビューを評価した後で，問合せを評価する．このとき，サーバ側に保持されている全ての日付の記事に関して，その見出しを連結するビュー定義を評価する必要がある．

ここで，マジックを適用することで必要のない日付の記事に関してはその見出しを連結せずに問合せ結果を返すことができる．すなわち，見出しに'Clinton'が出現しない日付けの記事は計算する必要がない．もし，該当する記事の日付が少ければ，マジックによる最適化手法を適用する価値がある．

図3に示した問合せにマジックによる最適化手法を適用した問合せを図5に示す．ここで，ビュー"Filter.xml"はマジックセットに相当し，ビュー"PartialResult.xml"がsupplementary関係に相当する．このsupplementary関係において，見出しに'Clinton'が出現するという条件を適用することで，もともと与えられた入力XML文書のうち見出しが束縛されるようなXML文書へと制限される．この制限されたXML文書のうち結合属性である日付を用いて自己結合する．この自己結合はいわゆるフィルタ結合と呼ばれるものである[SHP+96]．このようにオリジナルの問合せにおける検索条件をビュー定義へとpassingするのが，SIPS(Sideways Information Passing Strategy)と呼ばれる手法である．

[1]従来のマジックを拡張して*bcf* adornments関係を用いる，*ground magic-sets transformation*[MFPR96]もあるが，これについては本稿の範囲外である．

4. 結論と今後の課題

本稿では，入力要素を保持する集約の例としてXML文書を操作する問合せ言語を対象として，その最適化手法としてマジックによる書き換えを提案した．この手法を一般化することで，入れ子関係データベースやオブジェクト指向データベースに対する問合せの最適化へと適用可能である．

今後の課題としては，次の項目が挙げられる．

- マジックに基づく一般的なアルゴリズムの開発とその適用可能範囲の明確化．
- supplemnentary関係とマジック集合に相当するXML文書の生成と評価をサーバ側で行うか，クライアント側で行うか，ハイブリッドに行うかの指標の開発．

参考文献

[AHV95] Serge Abiteboul, Richard Hull, and Victor Vianu. *Foundations of Databases.* Addison-Wesley, 1995.

[BMSU86] F. Bancilhon, D. Maier, Y. Sagiv, and J. Ullman. Magic sets and other strange ways to implement logic programs. In *Proc. ACM SIGACT-SIGMOD Symp. on Principles of Database Sys.*, Boston, MA, 1986.

[BR87] C. Beeri and R. Ramakrishnan. On the power of magic. In *Proc. ACM SIGACT-SIGMOD Symp. on Principles of Database Sys.*, p. 269, San Diego, CA, March 1987.

[DFF+] Alin Deutsch, Mary F. Fernandez, Daniela Florescu, Alon Y. Levy, and Dan Suciu. A query language for xml.

[DFF+98] Alin Deutsch, Mary Fernandez, Daniela Florescu, Alon Levy, and Dan Suciu. XML-QL : A Query Language for XML, Aug 1998. http://www.w3.org/TR/NOTE-xml-ql/.

[MFPR90] I. S. Mumick, S. J. Finkelstein, Hamid Pirahesh, and Raghu Ramakrishnan. Magic is relevant. In *Proc. ACM SIGMOD International Conference on Management of Data*, pp. 247–258, May 1990.

[MFPR96] I. S. Mumick, S. J. Finkelstein, Hamid Pirahesh, and Raghu Ramakrishnan. Magic conditions. *ACM Transactions on Database Systems*, Vol. 21, No. 1, pp. 107–155, March 1996.

[SHP+96] Praveen Seshadri, Joseph M. Hellerstein, Hamid Pirahesh, T. Y. Cliff Leung, Raghu Ramakrishnan, Divesh Srivastava Peter J. Stuckey, and S. Sudarshan. Cost-based optimization for magic: Algebra and implementation. In *Proc. ACM SIGMOD International Conference on Management of Data*, pp. 435–446, June 1996.

[WWWC98] World Wide Web Consortium. eXtensible Markup Language (XML) 1.0. http://www.w3.org/TR/1998/REC-xml-19980210, February 1998. W3C Recommendation 10-February-1998.

ビュー定義のための問合せ

ビュー"PartialResult.xml"定義のための問合せ

```
WHERE <article>
        <date>$d</>
        <headline>$h</>
      </> ELEMENT_AS $a
      IN "www.newsserver./news.xml",
      $h CONTAIN 'Clinton'
CONSTRUCT $a
```

ビュー"Filter.xml"定義のための問合せ

```
WHERE <article>
        <date>$d</> ELEMENT_AS $ed
        <headline>$h</>
      </>
      IN "PartialResult.xml"
CONSTRUCT $ed
```

ビュー"LimitedHeadlinesByDate.xml"定義のための問合せ

```
WHERE <date>$d</>
      IN "Filter.xml",
      <article>
        <date>$d</>
        <headline></> ELEMENT_AS $h
      </>
      IN "www.newsserver/news.xml"
GROUP-BY $d
CONSTRUCT <headlines-by-date>
            <date>$d</>
            <headlines>
              cat($h)
            </>
          </>
```

Main Query Block

```
WHERE <headlines-by-date></>
        ELEMENT_AS $ans
CONSTRUCT $ans
```

図 5 マジックによる書き換え

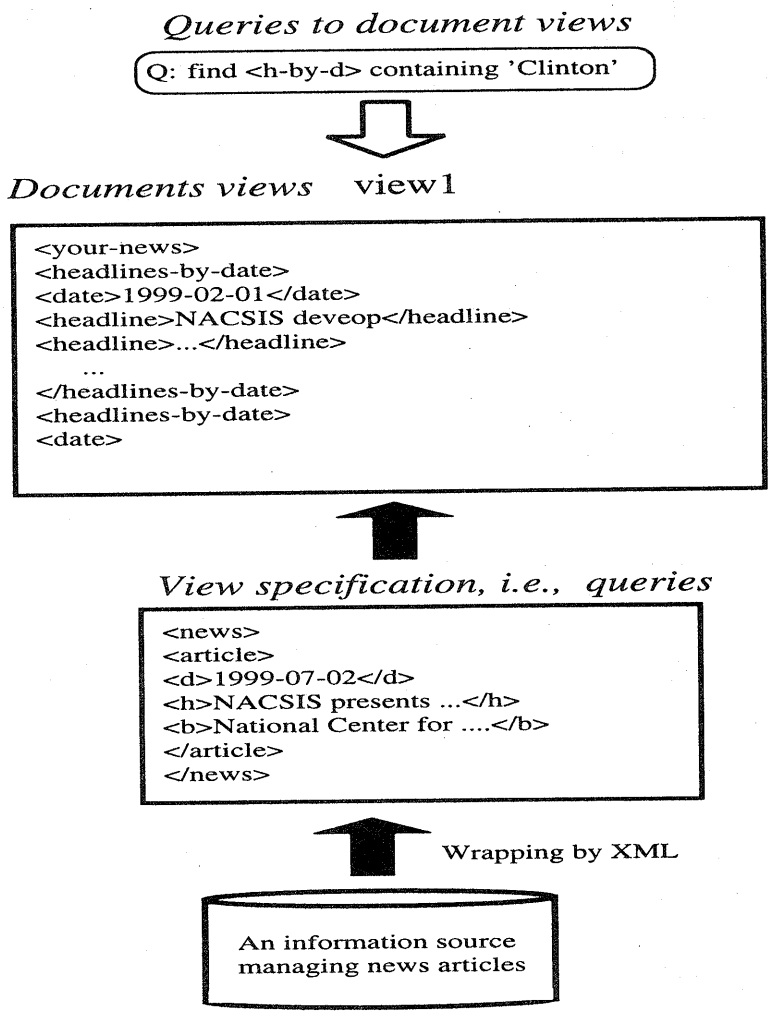

図 1 新聞記事に関する XML 文書ビュー

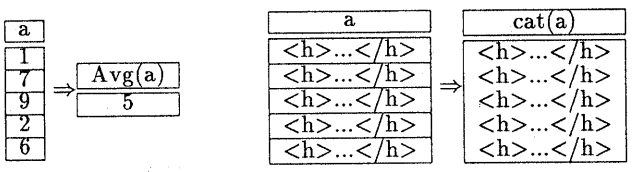

図 4 従来の集約と入力要素を保持する集約