

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5100820号
(P5100820)

(45) 発行日 平成24年12月19日(2012.12.19)

(24) 登録日 平成24年10月5日(2012.10.5)

(51) Int.Cl. F 1
G 0 6 F 17/30 (2006.01)
 G 0 6 F 17/30 1 4 0
 G 0 6 F 17/30 1 1 0 C
 G 0 6 F 17/30 3 3 0 B

請求項の数 11 (全 19 頁)

(21) 出願番号	特願2010-262178 (P2010-262178)	(73) 特許権者	000003078 株式会社東芝 東京都港区芝浦一丁目1番1号
(22) 出願日	平成22年11月25日(2010.11.25)	(73) 特許権者	301063496 東芝ソリューション株式会社 東京都港区芝浦一丁目1番1号
(65) 公開番号	特開2012-113524 (P2012-113524A)	(74) 代理人	100089118 弁理士 酒井 宏明
(43) 公開日	平成24年6月14日(2012.6.14)	(72) 発明者	金輪 拓也 東京都港区芝浦一丁目1番1号 東芝ソリューション株式会社内
審査請求日	平成22年11月25日(2010.11.25)	審査官	早川 学

最終頁に続く

(54) 【発明の名称】 問合せ式変換装置、方法およびプログラム

(57) 【特許請求の範囲】

【請求項1】

構造化データを蓄積するデータ集合に対する問合せ式の変換を行う問合せ式変換装置であって、

前記データ集合から定期的に前記構造化データを取得し、取得した前記構造化データを解析して、前記データ集合の構造情報を生成する第1の生成部と、

前記第1の生成部が生成した前記構造情報を前記データ集合のユニークアドレスと対応付けて記憶する第1の記憶部と、

問合せの対象となるデータ集合のユニークアドレスと、データの構造を条件にした問合せ式と、を入力する入力部と、

前記入力部が入力したデータ集合のユニークアドレスに対応する前記構造情報が前記第1の記憶部に記憶されていない場合に、該データ集合から指定された数の前記構造化データを取得し、取得した前記構造化データを解析して、該データ集合についての前記構造情報を生成する第2の生成部と、

前記入力部が入力した問合せ式を、前記入力部が入力したデータ集合のユニークアドレスに対応する前記構造情報に適合するように変換する変換部と、

前記変換部が変換した変換後問合せ式を出力する出力部と、を備えることを特徴とする問合せ式変換装置。

【請求項2】

前記第2の生成部は、前記入力部が入力した問合せ式に対して前記条件を緩和した問合せ

せ式を用いて前記データ集合から前記構造化データを取得することを特徴とする請求項1に記載の問合せ式変換装置。

【請求項3】

前記入力部が問合せの対象となるデータ集合のユニークアドレスとして複数のデータ集合のユニークアドレスを入力した場合に、入力された複数のデータ集合のユニークアドレスに対応する複数の前記構造情報を合成して、合成構造情報を生成する第3の生成部をさらに備え、

前記変換部は、前記入力部が入力した問合せ式を、前記合成構造情報に適合するように変換することを特徴とする請求項1または2に記載の問合せ式変換装置。

【請求項4】

構造化データを蓄積するデータ集合に対する問合せ式の変換を行う問合せ式変換装置であって、

前記データ集合から定期的に前記構造化データを取得し、取得した前記構造化データを解析して、前記データ集合の構造情報を生成する第1の生成部と、

前記第1の生成部が生成した前記構造情報を前記データ集合のユニークアドレスと対応付けて記憶する第1の記憶部と、

問合せの対象となるデータ集合のユニークアドレスと、データの構造を条件にした問合せ式と、を入力する入力部と、

前記入力部が入力した問合せ式を、前記入力部が入力したデータ集合のユニークアドレスに対応する前記構造情報に適合するように変換する変換部と、

前記変換部が変換した変換後問合せ式を出力する出力部と、を備え、

前記変換部は、前記構造情報に適合するように前記入力部が入力した問合せ式の前記条件を緩和した変換後問合せ式の候補を複数生成し、生成した複数の変換後問合せ式の候補に対して前記条件の緩和レベルに応じたスコア付けを行い、スコア値を評価基準として、複数の変換後問合せ式の候補の中から少なくとも1つの変換後問合せ式を選択することを特徴とする問合せ式変換装置。

【請求項5】

前記構造情報には、該構造情報を構成する各要素間の繋がり の発生確率を示す情報が付加されており、

前記変換部は、生成した複数の変換後問合せ式の候補に対して前記条件の緩和レベルと各要素間の繋がり の発生確率とに応じたスコア付けを行うことを特徴とする請求項4に記載の問合せ式変換装置。

【請求項6】

前記変換後問合せ式を、該変換後問合せ式に対して付与されたスコア値とともに記憶する第2の記憶部をさらに備え、

前記変換部は、生成した複数の変換後問合せ式の候補のうち、前記第2の記憶部に記憶された前記変換後問合せ式と合致する変換後問合せ式に対して、前記条件の緩和レベルに応じたスコア値と、前記第2の記憶部に記憶された前記変換後問合せ式に対して付与されたスコア値とに基づいたスコア付けを行うことを特徴とする請求項4に記載の問合せ式変換装置。

【請求項7】

前記出力部が出力した変換後問合せ式に対するユーザの評価を反映したスコア値を取得して前記第2の記憶部に格納する取得部をさらに備えることを特徴とする請求項6に記載の問合せ式変換装置。

【請求項8】

構造化データを蓄積するデータ集合に対する問合せ式の変換を行う問合せ式変換方法であって、

前記データ集合から定期的に前記構造化データを取得し、取得した前記構造化データを解析して、前記データ集合の構造情報を生成するステップと、

生成した前記構造情報を前記データ集合のユニークアドレスと対応付けてメモリに記憶

10

20

30

40

50

するステップと、

問合せの対象となるデータ集合のユニークアドレスと、データの構造を条件にした問合せ式と、を入力するステップと、

入力したデータ集合のユニークアドレスに対応する前記構造情報が前記メモリに記憶されていない場合に、該データ集合から指定された数の前記構造化データを取得し、取得した前記構造化データを解析して、該データ集合についての前記構造情報を生成するステップと、

入力した問合せ式を、入力したデータ集合のユニークアドレスに対応する前記構造情報に適合するように変換するステップと、

変換した変換後問合せ式を出力するステップと、を含むことを特徴とする問合せ式変換方法。

10

【請求項 9】

構造化データを蓄積するデータ集合に対する問合せ式の変換を行う問合せ式変換方法であって、

前記データ集合から定期的に前記構造化データを取得し、取得した前記構造化データを解析して、前記データ集合の構造情報を生成するステップと、

生成した前記構造情報を前記データ集合のユニークアドレスと対応付けてメモリに記憶するステップと、

問合せの対象となるデータ集合のユニークアドレスと、データの構造を条件にした問合せ式と、を入力するステップと、

20

入力した問合せ式を、入力したデータ集合のユニークアドレスに対応する前記構造情報に適合するように変換するステップと、

変換した変換後問合せ式を出力するステップと、を含み、

前記入力した問合せ式を変換するステップでは、前記構造情報に適合するように前記入力した問合せ式の前記条件を緩和した変換後問合せ式の候補を複数生成し、生成した複数の変換後問合せ式の候補に対して前記条件の緩和レベルに応じたスコア付けを行い、スコア値を評価基準として、複数の変換後問合せ式の候補の中から少なくとも1つの変換後問合せ式を選択することを特徴とする問合せ式変換方法。

【請求項 10】

構造化データを蓄積するデータ集合に対する問合せ式の変換を行う問合せ式変換プログラムであって、

30

コンピュータに、

前記データ集合から定期的に前記構造化データを取得し、取得した前記構造化データを解析して、前記データ集合の構造情報を生成する機能と、

生成した前記構造情報を前記データ集合のユニークアドレスと対応付けてメモリに記憶する機能と、

問合せの対象となるデータ集合のユニークアドレスと、データの構造を条件にした問合せ式と、を入力する機能と、

入力したデータ集合のユニークアドレスに対応する前記構造情報が前記メモリに記憶されていない場合に、該データ集合から指定された数の前記構造化データを取得し、取得した前記構造化データを解析して、該データ集合についての前記構造情報を生成する機能と

40

、
入力した問合せ式を、入力したデータ集合のユニークアドレスに対応する前記構造情報に適合するように変換する機能と、

変換した変換後問合せ式を出力する機能と、を実現させることを特徴とする問合せ式変換プログラム。

【請求項 11】

構造化データを蓄積するデータ集合に対する問合せ式の変換を行う問合せ式変換プログラムであって、

コンピュータに、

50

前記データ集合から定期的に前記構造化データを取得し、取得した前記構造化データを解析して、前記データ集合の構造情報を生成する機能と、

生成した前記構造情報を前記データ集合のユニークアドレスと対応付けてメモリに記憶する機能と、

問合せの対象となるデータ集合のユニークアドレスと、データの構造を条件にした問合せ式と、を入力する機能と、

入力した問合せ式を、入力したデータ集合のユニークアドレスに対応する前記構造情報に適合するように変換する機能と、

変換した変換後問合せ式を出力する機能と、を実現させ、

前記入力した問合せ式を変換する機能は、前記構造情報に適合するように前記入力した問合せ式の前記条件を緩和した変換後問合せ式の候補を複数生成し、生成した複数の変換後問合せ式の候補に対して前記条件の緩和レベルに応じたスコア付けを行い、スコア値を評価基準として、複数の変換後問合せ式の候補の中から少なくとも1つの変換後問合せ式を選択することを特徴とする問合せ式変換プログラム。

10

【発明の詳細な説明】

【技術分野】

【0001】

本発明の実施の形態は、問合せ式変換装置、方法およびプログラムに関するものである。

【背景技術】

20

【0002】

汎用的なデータ記述言語であるXML (Extensible Markup Language) は、定型データと非定型データとを1つの文書中に混在させた構造を取ることが可能で、コンピュータが処理しやすく、かつ、人間にとっても可読性が高いという特徴を持つ。このため、XMLは各種の電子化データに共通のデータフォーマットとして広く普及している。XMLで記述された電子化データ(以下、XMLデータという。)は、タグ付け(マークアップ)された要素をツリー構造で表現した構造化データである。

【0003】

XMLデータが大量に出回ることによって、近年では、XMLデータを蓄積して適切に管理するのに適したXMLデータベースが開発されている。また、標準問合せ言語であるXQueryで記述した問合せ式を用い、ネットワーク上に分散された複数のXMLデータベース(XMLデータのデータ集合)をあたかも1つのデータベースとして検索を行う統合検索の技術開発も進められている。

30

【0004】

統合検索において問題となるのは、統合対象となる複数のXMLデータベースにおけるスキーマ(XMLデータのデータ構造)の違いである。XMLデータベースは、スキーマ定義を不要としているため、多種多様なXMLデータに対応できる柔軟性と拡張性に優れるが、蓄積されているXMLデータのデータ構造がデータベースごとに異なる場合がある。このため、統合対象となる複数のXMLデータベースに対して共通の問合せ式を実行したとしても、あるXMLデータベースでは検索条件に適合するXMLデータがヒットしたが、別のXMLデータベースでは、本来意味的には適合するはずのXMLデータが、そのデータ構造の違いからヒットしないという結果になることがあり得る。特に、問合せ言語としてXQueryを用いた場合は、構造や語彙を条件とした厳格なマッチングが基本となるため、この傾向は強い。

40

【0005】

このような複数のデータベース間のスキーマの違いを吸収し、共通の問合せ式で統合検索を行えるようにする方法として、共通の問合せ式を複数のデータベースの各々に適合した問合せ式に変換するマスタサーバをネットワーク上に設けるという方法が知られている。一般的には、統合対象となるデータベースごとに問合せ式変換のルールを事前に定義してマスタサーバに格納しておき、マスタサーバに対して共通の問合せ式を入力すると、マ

50

スタサーバで共通の問合せ式を統合対象となる複数のデータベースごとのスキーマに合わせて変換し、各データベースに対する問合せを実行するようにしている。

【先行技術文献】

【特許文献】

【0006】

【特許文献1】特許第3671765号公報

【特許文献2】特開2007-257083号公報

【発明の概要】

【発明が解決しようとする課題】

【0007】

しかしながら、問合せ式を変換する従来の方法では、統合対象となる各データベースのスキーマに従った変換ルールを事前に定義しておくことが必要であり、スキーマが未知のデータベースには対応できないという問題があった。

【課題を解決するための手段】

【0008】

実施の形態の問合せ式変換装置は、構造化データを蓄積するデータ集合に対する問合せ式の変換を行う。実施の形態の問合せ式変換装置は、第1の生成部と、第1の記憶部と、入力部と、第2の生成部と、変換部と、出力部と、を備えている。第1の生成部は、前記データ集合から定期的に前記構造化データを取得し、取得した前記構造化データを解析して、前記データ集合の構造情報を生成する。第1の記憶部は、前記第1の生成部が生成した前記構造情報を前記データ集合のユニークアドレスと対応付けて記憶する。入力部は、問合せの対象となるデータ集合のユニークアドレスと、データの構造を条件にした問合せ式と、を入力する。第2の生成部は、前記入力部が入力したデータ集合のユニークアドレスに対応する前記構造情報が前記第1の記憶部に記憶されていない場合に、該データ集合から指定された数の前記構造化データを取得し、取得した前記構造化データを解析して、該データ集合についての前記構造情報を生成する。変換部は、前記入力部が入力した問合せ式を、前記入力部が入力したデータ集合のユニークアドレスに対応する前記構造情報に適合するように変換する。出力部は、前記変換部が変換した変換後問合せ式を出力する。

【図面の簡単な説明】

【0009】

【図1】「特許」に関する情報をXMLで記述した例を示す図。

【図2】構造データガイドを説明する図。

【図3】合成構造データガイドを生成する様子を概念的に示す図。

【図4】統合検索システムの全体構成を概略的に示すシステム構成図。

【図5】第1の実施形態にかかる問合せ式変換装置の機能構成を示す機能ブロック図。

【図6】元問合せ式を緩和した問合せ式を用いて未知のDBサーバについての構造データガイドを生成する処理の一例を説明する図。

【図7】変換部による処理の具体的な一例を説明する図。

【図8】各要素間の繋がりや発生確率の情報が付加された合成構造データガイドの一例を示す図。

【図9】問合せ式変換処理部によって実行される一連の処理の流れを示したフローチャート。

【図10】第2の実施形態にかかる問合せ式変換装置の機能構成を示す機能ブロック図。

【図11】変換部による処理の具体的な一例を説明する図。

【発明を実施するための形態】

【0010】

(第1の実施形態)

本実施形態にかかる問合せ式変換装置は、ネットワーク上のサーバとして構成され、ネットワーク上のクライアント端末から入力された問合せ式を、問合せの対象となるネットワーク上のデータ集合である各データベースに適合した問合せ式に変換してクライアント

10

20

30

40

50

端末に出力するものである。ここで、ネットワーク上の各データベースは、構造化データであるXMLデータを格納しているものとする。また、クライアント端末から入力される問合せ式は、データの構造を条件にした問合せ式、具体的にはXQueryで記述された問合せ式であるものとする。

【0011】

まず、問合せの対象となる各データベースに蓄積されたデータの共通のフォーマットとして用いるXMLについて説明する。

【0012】

図1は、「特許」に関する情報をXMLで記述した例である。XMLでは、文書構造の表現にタグ(要素)が用いられる。タグには、開始タグと終了タグがあり、文書構造情報の構成要素を開始タグと終了タグで囲むことにより、文書中の文字列(テキスト)区切りと、そのテキストが構造上どの構成要素に属するのかを明確に記述することができる。

10

【0013】

ここで、開始タグとは、要素名称を記号「<」、「>」で閉じたものであり、終了タグとは、要素名称を記号「</」と「>」で閉じたものである。タグに続く構成要素の内容が、テキスト(文字列)または子供の構成要素の繰り返しである。また、開始タグには、「<要素名称 属性="属性値">」などのように属性情報を設定することができる。「<特許DB></特許DB>」のようにテキストを含まない構成要素は、簡易記法として「<特許DB/>」のように表すこともできる。

【0014】

20

図1に示した文書は、「特許」タグから始まる要素を文書ルート(根)とし、その子要素として「タイトル」、「発明者リスト」、「効果」、「キーワードリスト」タグから始まる要素集合が存在し、また、「タイトル」タグから始まる要素には「構造化文書検索装置」といった、1つのテキスト(文字列)が存在することを示している。

【0015】

次に、問合せの対象となる各データベースにおいて特徴的な構造情報である構造データガイドについて説明する。構造データガイドは、図2に示すように、同一構造において同一タグ名を持つ要素を集約して1つのデータ構造として表したものであり、XMLの階層構造を抽象化して表現したものである。図2の例の場合は、発明者タグが同一構造で複数回存在しているが、これらが集約された形で構造データガイドが生成されている。構造データガイドは文書単位で生成されるが、複数の文書の場合は、まず、その文書に対する構造データガイドを求め、その情報と、これまでに生成された構造データガイドを重ね合わせることで、新たに構造データガイドを更新していくことになる。つまり、1つのデータベースから複数のXMLデータを取り出し、その構造を解析して以上のような構造データガイドを逐次生成し、更新していくことにより、そのデータベースにおいて特徴的な構造情報である構造データガイドを得ることができる。この構造データガイドは、データベースにおけるスキーマを仮想的に表したものとなる。

30

【0016】

また、以上のように各データベースごとに生成された構造データガイドを複数のデータベース間で合成すれば、複数のデータベースを統合した場合(統合データベース)の仮想的なスキーマとなる合成構造データガイドを得ることができる。複数の構造データガイドを合成して合成構造データガイドを生成する様子を図3に概念的に示す。図3では、4つのデータベースの構造データガイドを合成して1つの合成構造データガイドを生成した例を簡略化して示している。図3に示すように、合成構造データガイドは、元の構造データガイドにおいて共通する構造を持つ部分を1つのノードとして表現している。

40

【0017】

図4は、本実施形態にかかる問合せ式変換装置を用いた統合検索システムの全体構成を概略的に示すシステム構成図である。この統合検索システムは、本実施形態にかかる問合せ式変換装置10と、問合せ式変換装置10によるサービスを受けるクライアント端末20と、複数のDBサーバ(XMLデータベース)30とを備え、これらの各構成要素がネ

50

ットワークにより接続された構成である。

【 0 0 1 8 】

クライアント端末 2 0 は、問合せの対象となる複数の DB サーバ 3 0 それぞれのネットワーク上のユニークアドレスである複数の URL と、 X Q u e r y で記述された問合せ式（元問合せ式）とを、問合せ式変換装置 1 0 に対して出力する。また、クライアント端末 2 0 は、問合せ式変換装置 1 0 によって元問合せ式を変換することで得られた変換後問合せ式を、問合せの対象となる複数の DB サーバ 3 0 に対してそれぞれ出力し、変換後問合せ式に対する応答として各 DB サーバ 3 0 から出力される検索結果を取得する。

【 0 0 1 9 】

ここで、クライアント端末 2 0 は、ユーザが使用するパーソナルコンピュータなどの端末（ユーザ端末）であってもよいし、複数の DB サーバ 3 0 を取りまとめるコーディネータサーバであってもよい。複数の DB サーバ 3 0 を取りまとめるコーディネータサーバをネットワーク上に設けた場合、コーディネータサーバは、問合せの対象となる複数の DB サーバ 3 0 それぞれの URL と元問合せ式とをユーザ端末から入力して、問合せ式変換装置 1 0 に対して出力する。そして、問合せ式変換装置 1 0 で変換された変換後問合せ式を問合せの対象となる複数の DB サーバ 3 0 に対して出力し、各 DB サーバ 3 0 から検索結果を取得してユーザ端末に出力する。このように、コーディネータサーバは、問合せ式変換装置 1 0 との関係において、クライアント端末 2 0 となる。

【 0 0 2 0 】

DB サーバ 3 0 は、XML データを格納したデータベースを管理するサーバであり、入力した問合せ式を用いてデータベースを検索し、その検索結果を問合せ式への応答として出力する。ネットワーク上の各 DB サーバ 3 0 は、ネットワーク上のユニークアドレスである URL により識別される。ここで、統合検索システムを構成する複数の DB サーバ 3 0 には、URL が事前に問合せ式変換装置 1 0 に登録されている DB サーバ（つまり、問合せ式変換装置 1 0 にとって既知の DB サーバ）と、URL が問合せ式変換装置 1 0 に登録されていない DB サーバ（つまり、問合せ式変換装置 1 0 にとって未知の DB サーバ）とが含まれる。

【 0 0 2 1 】

図 5 は、本実施形態にかかる問合せ式変換装置 1 0 の詳細な機能構成を示す機能ブロック図である。問合せ式変換装置 1 0 は、第 1 の生成部 1 1 と、第 1 の記憶部 1 2 と、入力部 1 3 と、出力部 1 4 と、問合せ式変換処理部 1 5 とを備える。問合せ式変換処理部 1 5 には、URL 照合部 1 5 1 と、第 2 の生成部 1 5 2 と、第 3 の生成部 1 5 3 と、変換部 1 5 4 とが設けられている。

【 0 0 2 2 】

第 1 の生成部 1 1 は、URL が事前に登録されている既知の DB サーバ 3 0 に対して定期的にクローリングを実施して、既知の DB サーバ 3 0 の構造情報である構造データガイドを生成する。具体的には、第 1 の生成部 1 1 は、URL が登録されているすべての DB サーバ 3 0 に対して、例えば、X Q u e r y の単純パス表現である / * のように、必ずルートから全件マッチさせる問合せ式を定期的に行い、各 DB サーバ 3 0 からデータベースに格納されている XML データを各々取得する。そして、第 1 の生成部 1 1 は、取得した XML データを解析することで、各 DB サーバ 3 0 が管理するデータベースの特徴的な構造情報（仮想的なスキーマ）である構造データガイド（図 2 参照）を定期的に生成し、生成した構造データガイドを各 DB サーバ 3 0 の URL に対応付けて第 1 の記憶部 1 2 に格納する。なお、DB サーバ 3 0 が事前に外部に対して構造データガイドを公開している場合には、第 1 の生成部 1 1 は、DB サーバ 3 0 から XML データを取得して構造データガイドを生成する代わりに、この公開されている構造データガイドを取得すればよい。

【 0 0 2 3 】

第 1 の記憶部 1 2 は、第 1 の生成部 1 1 による定期的なクローリングによって生成された既知の DB サーバ 3 0 についての構造データガイドを、その DB サーバ 3 0 のユニークアドレスである URL と対応付けて記憶するメモリである。第 1 の生成部 1 1 による既知

10

20

30

40

50

のDBサーバ30に対するクローリングは上記のように定期的実施されるため、各URLに対応する構造データガイドは定期的更新され、常に最新の構造データガイドが第1の記憶部12に記憶されることになる。

【0024】

入力部13は、クライアント端末20から出力された問合せ対象となる複数のDBサーバ30それぞれのURL(URL集合)と、XQueryで記述された元問合せ式とを入力する。そして、入力部13は、入力したURL集合をURL照合部151に受け渡すとともに、入力した元問合せ式を変換部154に受け渡す。

【0025】

出力部14は、変換部154によって元問合せ式が変換されることで得られた変換後問合せ式を取得し、この変換後問合せ式をクライアント端末20に対して出力する。

10

【0026】

URL照合部151は、入力部13が入力したURL集合に含まれる各URLに対応する構造データガイドが第1の記憶部12に記憶されているか否かを確認する。そして、URL照合部151は、入力部13が入力したURL集合に、第1の記憶部12に構造データガイドが記憶されていないURL、つまり、未知のDBサーバ30のURLが含まれている場合には、第2の生成部152に対して、未知のDBサーバ30についての構造データガイドを生成することを要求する。

【0027】

第2の生成部152は、URL照合部151からの要求に応じて、未知のDBサーバ30についての構造データガイドを生成する処理を実行する。具体的には、第2の生成部152は、未知のDBサーバ30に対する問合せ式を生成し、この問合せ式に対する応答として未知のDBサーバ30から得られたN件のXMLデータを解析し、未知のDBサーバ30についての構造データガイドを生成する。なお、未知のDBサーバ30から取得するXMLデータの件数Nが大きいほど、生成する構造データガイドの精度(つまり、スキーマの再現性)は向上するが、構造データガイドの生成に要する処理時間が長くなる。したがって、未知のDBサーバ30から取得するXMLデータの件数Nは、要求される構造データガイドの精度と許容される処理時間との兼ね合いから、最適な件数を事前に設定しておくことが望ましい。

20

【0028】

未知のDBサーバ30に対する問合せ式としては、第1の生成部11が実行する問合せ式と同様、XQueryの単純パス表現である/*のように、必ずルートから全件マッチさせる問合せ式を用いるようにしてもよいが、問合せの効率性や精度を高めるために、入力部13が入力した元問合せ式の条件を緩和した問合せ式を用いることが望ましい。

30

【0029】

図6は、入力部13が入力した元問合せ式の条件を緩和した問合せ式を用いて、未知のDBサーバ30についての構造データガイドを生成する処理の一例を説明する説明図である。本例では、入力部13が入力した元問合せ式が、図6(a)に示すように、XQueryで記述された問合せ式Aであるものとする。XQueryで記述された問合せ式のようにデータの構造を条件にした問合せ式は、一般的に、グラフ構造に変形して処理される。図6(a)に示す問合せ式Aは、図6(b)に示すような問合せグラフに変形されて処理される。この問合せグラフは、タグ名をノードとし、アークにはパス制約などの制約情報を記述している。問合せ処理は、この問合せグラフの制約を満たすXMLデータを、未知のDBサーバ30が管理するデータベース(URL1)から抽出する処理となる。

40

【0030】

本例では、未知のDBサーバ30(URL1)が管理するデータベースには、図6(c)に示すように、文書1、文書2、文書3の3つのXMLデータが格納されているものとし、これらのXMLデータのうち、問合せグラフの制約を満たすXMLデータを抽出するものとする。ここで、元の問合せ式Aのままでは、構造の違いにより、URL1上では1件もヒットしないことになる。そこで、図6(d)に示すように、問合せ式Aの条件を緩

50

和した緩和問合せ式 A' を生成する。この緩和問合せ式 A' を生成する方法に関してはさまざまな方法が考えられるが、ここでは、テキストの抽出条件だけを残し、構造に関する制約条件を完全に緩和する方法を適用する。緩和方法の他の例としては、例えば、タグ名のシソーラス情報を用いた緩和や、発生確率が高い構造を緩和しない、などが考えられる。

【 0 0 3 1 】

図 6 (d) に示す緩和問合せ式 A' を未知の DB サーバ 3 0 (URL 1) に対して実行すると、文書 1 および文書 2 の XML データを得ることができる。そして、これらの XML データを解析することで、図 6 (e) のように、未知の DB サーバ 3 0 (URL 1) の仮想的なスキーマとなる構造データガイドを生成することができる。

10

【 0 0 3 2 】

なお、本例では簡単のために、未知の DB サーバ 3 0 (URL 1) が管理するデータベースに 3 つの XML データのみが格納され、そのうちの 2 件がヒットした場合を例示しているが、実際にはより多数の XML データがデータベースに格納されているのが一般的であり、ヒット件数も多くなる。したがって、この場合も処理時間を高速化するために、指定された (先頭から) N 件の XML データを取得した段階で処理を打ち切る。この場合は、X Query の単純パス表現である / * を問合せ式として用いる場合と異なり、解空間をある程度絞り込んでいるため、単純に / * で検索するよりも、少ないデータ件数で高い精度を持つ構造データガイドを生成できることになる。また、未知の DB サーバ 3 0 が事前に構造データガイドを公開している場合には、第 1 の生成部 1 1 による処理と同様、未知の DB サーバ 3 0 から N 件の XML データを取得して構造データガイドを生成する代わりに、この公開されている構造データガイドを取得すればよい。

20

【 0 0 3 3 】

第 3 の生成部 1 5 3 は、入力部 1 3 が入力した URL 集合に含まれる各 URL のうちで、構造データガイドが第 1 の記憶部 1 2 に記憶されているものについては、該当する構造データガイドを第 1 の記憶部 1 2 からそれぞれ取り出す。また、構造データガイドが第 1 の記憶部 1 2 に記憶されていないものについては、第 2 の生成部 1 5 2 により生成された構造データガイドを取得する。そして、第 3 の生成部 1 5 3 は、これら入力部 1 3 が入力した URL 集合に含まれる各 URL に対応する複数の構造データガイドを、図 3 に示したように合成することによって、合成構造データガイドを生成する。

30

【 0 0 3 4 】

このように生成された合成構造データガイドは、問合せ対象となる複数の DB サーバ 3 0 を統合した統合データベースの仮想的なスキーマとなるものであり、仮に問合せ対象となる DB サーバ 3 0 の一部の構造情報が欠落していたとしても、類似の構造情報を持つ他の DB サーバ 3 0 の情報によって補完されるため、スキーマの再現性が高い。また、合成構造データガイドは、DB サーバ 3 0 に対する問合せを実行する際に、第 1 の記憶部 1 2 に格納されている構造データガイドや第 2 の生成部 1 5 2 により生成される構造データガイドを用いて生成される (つまり、動的に生成される) ため、データベース内の XML データの追加や削除などによって生じる DB サーバ 3 0 のスキーマ変化にも対応し、時々刻々と変化する統合データベースのスキーマを精度よく再現した情報となる。

40

【 0 0 3 5 】

変換部 1 5 4 は、入力部 1 3 が入力した元問合せ式を、第 3 の生成部 1 5 3 により生成された合成構造データガイドに適合するように変換して、変換後問合せ式を生成する。上述したように、第 3 の生成部 1 5 3 により生成された合成構造データガイドは、統合データベースの仮想的なスキーマを表した情報であり、問合せ対象となる複数の DB サーバ 3 0 上で発生確率が高い構造、あるいは、完全にマッチしなくても、類似構造をとりやすい構造を表している。変換部 1 5 4 は、入力部 1 3 が入力した元問合せ式に対して、もしこのような構造データガイドを持つデータベースが存在した場合に、元問合せ式に対してどのような変換を行えばマッチングする可能性が高まるか、という観点から元問合せ式を変形する。

50

【 0 0 3 6 】

変換部 1 5 4 による処理の具体的な一例について図 7 を用いて説明する。図 7 に示す例は、入力部 1 3 が入力した元問合せ式が / A / B / C であった場合の例であり、合成構造データガイドに対してマッチング処理を行う例である。まず、/ A / B / C の元問合せ式そのものをマッチング条件とした場合に、合成構造データガイド上で直接マッチするパスは存在しないので、この元問合せ式を変形する。この方法は組み合わせ最適化問題であるが、現実的にはヒューリスティックな方法により解決される。例えば、ある制約条件を緩和するごとにスコア値を下げていき、スコア値が最も高くなるような組合せを決定する、といった方法などが考えられる。

【 0 0 3 7 】

ここでは、制約条件の緩和の例として、(1) / を / / に緩和する場合、(2) 未知要素を新規追加する場合、(3) 同一構造内の要素を O R で繋ぐ場合、を想定し、(1) の制約緩和コストを 3、(2) の制約緩和コストを 7、(3) の制約緩和コストを 5 としている。また、元問合せ式の初期スコアを 1 0 0 点としている。そして、上記 (1) ~ (3) に示す制約条件の緩和を行うことで合成構造データガイドとマッチングすることになった変換後問合せ式の候補をそれぞれ生成し、それぞれの制約緩和コストを元問合せ式の初期スコアから減算して、変換後問合せ式の候補のスコア値を求める。そして、最もスコア値の高い変換後問合せ式の候補を選択して、変換後問合せ式とする。

【 0 0 3 8 】

図 7 の例の場合、元問合せ式に対して (1) の制約緩和ルールを適用した / A / / B / C が、変換後問合せ式の候補の 1 つとして挙げられる。この変換後問合せ式の候補 / A / / B / C のスコア値は、 $100 - 3 = 97$ となる。また、元問合せ式に対して (2) の制約緩和ルールを適用した / A / G / B / C も、構造データガイド上でマッチするパスが存在するため、変換後問合せ式の候補の 1 つとなる。この変換後問合せ式の候補 / A / G / B / C のスコア値は、 $100 - 7 = 93$ となる。また、元問合せ式に対して (3) の制約緩和ルールを適用した / A / (B | C | D) / C も、変換後問合せ式の候補の 1 つとなる。この変換後問合せ式の候補 / A / (B | C | D) / C のスコア値は、 $100 - 5 = 95$ となる。

【 0 0 3 9 】

以上の 3 つの変換後問合せ式の候補の中では、元問合せ式に対して (1) の制約緩和ルールを適用した / A / / B / C のスコア値が最も高い。したがって、変換部 1 5 4 は、これらの候補の中から変換後問合せ式として / A / / B / C を選択し、入力部 1 3 が入力した元問合せ式 / A / B / C を / A / / B / C に変換して、出力部 1 4 に受け渡す。これにより、出力部 1 4 からクライアント端末 2 0 に対して、問合せ対象となる複数の D B サーバ 3 0 についての合成構造データガイドに適合するかたちに変換された変換後問合せ式である / A / / B / C が出力される。

【 0 0 4 0 】

ところで、合成構造データガイドやその元となる個々の構造データガイドは、基本的に重ねあわせによって生成される。したがって、合成構造データガイドや個々の構造データガイドには、例えば図 8 に示すように、各要素間の繋がり の発生確率を情報として付加することができる。そして、合成構造データガイドや個々の構造データガイドに付加された各要素間の繋がり の発生確率を示す情報は、変換部 1 5 4 での変換後問合せ式の候補に対するスコア付けに利用することができる。例えば、図 8 に例示するように、A という要素の直下には 0 . 2 の確率で B という要素が存在し、0 . 3 の確率で C という要素が存在することが分かるとすれば、変換後問合せ式の候補のスコア値として、/ A / B を含む変換後問合せ式の候補よりも、/ A / C を含む変換後問合せ式の候補のスコア値を高くする、ということが考えられる。

【 0 0 4 1 】

次に、本実施形態にかかる問合せ式変換装置 1 0 の動作について、図 9 を参照して説明する。図 9 は、入力部 1 3 がクライアント端末 2 0 からの URL 集合と元問合せ式を入力

10

20

30

40

50

したときに、問合せ式変換処理部 15 によって実行される一連の処理の流れを示したフローチャートである。なお、本実施形態にかかる問合せ式変換装置 10 では、図 9 に示す問合せ式変換処理部 15 による処理とは別ルーチンで、第 1 の生成部 11 による既知の DB サーバ 30 に対する定期的なクローリングが実行され、第 1 の記憶部 12 には、入力部 13 が URL 集合および元問合せ式を入力した時点での既知の DB サーバ 30 についての最新の構造データガイドが記憶されているものとする。

【 0042 】

図 9 のフローチャートで示す処理が開始されると、まず、クライアント端末 20 から入力された URL 集合が、問合せ式変換処理部 15 の URL 照合部 151 に取り込まれる (ステップ S101)。URL 照合部 151 は、取り込んだ URL 集合に含まれる各 URL をキーとして第 1 の記憶部 12 を検索し、第 1 の記憶部 12 に構造データガイドが記憶されていない URL が存在するか否かを判定する (ステップ S102)。

10

【 0043 】

ここで、第 1 の記憶部 12 に構造データガイドが記憶されていない URL が存在する場合 (ステップ S102: Yes)、つまり、問合せ式変換装置 10 にとって未知の DB サーバ 30 の URL が URL 集合に含まれている場合には、第 2 の生成部 152 が、未知の DB サーバ 30 にアクセスして、この未知の DB サーバ 30 から N 件の XML データを抽出する (ステップ S103)。そして、第 2 の生成部 152 は、抽出した N 件の XML データを解析して、未知の DB サーバ 30 についての構造データガイドを生成する (ステップ S104)。なお、未知の DB サーバ 30 が構造データガイドを公開している場合には、第 2 の生成部 152 は、この公開されている構造データガイドを取得する。

20

【 0044 】

一方、第 1 の記憶部 12 に構造データガイドが記憶されていない URL が存在しない場合 (ステップ S102: No)、つまり、URL 集合に含まれるすべての URL が問合せ式変換装置 10 に登録された既知の DB サーバ 30 の URL であり、第 1 の生成部 11 による定期的なクローリングによって構造データガイドが生成されている場合には、ステップ S105 へと処理を移行する。

【 0045 】

次に、第 3 の生成部 153 が、第 1 の記憶部 12 に記憶されている既知の DB サーバ 30 についての構造データガイドと、未知の DB サーバ 30 が存在する場合には第 2 の生成部 152 により生成された構造データガイドとを取り込み、URL 集合に含まれるすべての URL に対応する構造データガイドを合成して、問合せの対象となるすべての DB サーバ 30 を統合した統合データベースの仮想的なスキーマとなる合成構造データガイドを生成する (ステップ S105)。

30

【 0046 】

次に、変換部 154 が、クライアント端末 20 から入力された元問合せ式を取り込み、この元問合せ式の制約条件を緩和しながら合成構造データガイドとのマッチングを行って、制約条件を緩和した複数の変換後問合せ式の候補を生成する (ステップ S106)。このとき、変換部 154 は、生成した複数の変換後問合せ式の候補に対して、制約条件の緩和レベルに応じたスコア付けを行う。そして、変換部 154 は、生成した複数の変換後問合せ式の候補のうちで、スコア値が最も高い候補を、変換後問合せ式として選択する (ステップ S107)。この変換後問合せ式は、変換部 154 から出力部 14 へと受け渡され、出力部 14 からクライアント端末 20 に対して出力される (ステップ S108)。

40

【 0047 】

なお、図 9 に示すフローチャートは、処理の途中で第 1 の記憶部 12 に記憶されている構造データガイドに変化がない場合を想定しているが、第 1 の生成部 11 によるクローリングは別ルーチンで実行されるため、処理の途中で、URL 集合に含まれるいずれかの URL に対応する構造データガイドが更新される場合もある。このような場合には、第 3 の生成部 153 は、その時点で、更新された構造データガイドを用いて合成構造データガイドを生成することが望ましい。これにより、問合せの対象となる DB サーバ 30 の最新の

50

状態に対応した合成構造データガイドを用いて元問合せ式を変換することができ、変換の精度が向上する。

【0048】

また、以上説明した例では、変換部154が元問合せ式を1つの変換後問合せ式に変換するようにしているが、URL集合に含まれるURLごとに変換後問合せ式を生成することも可能である。この場合、変換部154は、元問合せ式の制約条件を緩和して合成構造データガイドとのマッチングを計算する際に、それぞれのURLに対応する構造データガイドとのマッチングも考慮して、例えば、対応する構造データガイドにマッチする変換後問合せ式の候補のスコアを高くするといった手法により、URLごとに異なった変換後問合せ式を選択すればよい。

10

【0049】

以上、具体的な例を挙げながら詳細に説明したように、本実施形態にかかる問合せ式変換装置10は、既知のDBサーバ30については定期的なクローリングによって仮想的なスキーマとなる構造データガイドを生成して第1の記憶部11に格納しておき、未知のDBサーバ30については、元問合せ式の変換を行う際に仮想的なスキーマとなる構造データガイドを生成する。そして、これら構造データガイドを用いて、クライアント端末20から入力した元問合せ式を変換し、変換後問合せ式をクライアント端末20に出力する。したがって、ネットワーク上の各DBサーバ30についてのスキーマに従った変換ルールを事前に定義しておくことなく、問合せの対象となるDBサーバ30に適合するように問合せ式を適切に変換することができ、また、スキーマが未知のDBサーバ30にも対応することができる。

20

【0050】

また、本実施形態にかかる問合せ式変換装置10は、問合せの対象となるDBサーバ30が複数の場合には、これら複数のDBサーバ30についての構造データガイドを合成して、複数のDBサーバ30を統合した統合データベースの仮想的なスキーマとなる合成構造データガイドを生成する。そして、クライアント端末20から入力した元問合せ式をこの合成構造データガイドに適合するように変換する。したがって、仮に問合せの対象となるDBサーバ30の一部の構造情報が欠落していたとしても、類似の構造情報を持つ他のDBサーバ30の情報によって補完して、問合せの対象となるすべてのDBサーバ30に対して有効な変換後問合せ式を得ることができ、統合検索システムにおける検索の精度向上を実現することができる。

30

【0051】

(第2の実施形態)

次に、第2の実施形態の問合せ式変換装置について説明する。本実施形態にかかる問合せ式変換装置には、過去の問合せ式の変換処理の履歴を利用して、複数の変換後問合せ式の候補に対するスコア付けを行うための機能が付加されている。すなわち、本実施形態にかかる問合せ式変換装置では、過去の変換処理によって得た元問合せ式に対する変換後問合せ式を、その変換後問合せ式に対して付与されたスコア値とともに記憶しておき、その後の変換処理において複数の変換後問合せ式の候補に対してスコア付けを行う際に、記憶したスコア値も加味してスコア付けを行うようにしている。

40

【0052】

図10は、本実施形態にかかる問合せ式変換装置10'の機能構成を示す機能ブロック図である。本実施形態にかかる問合せ式変換装置10'においては、上述した第1の実施形態にかかる問合せ式変換装置10の構成(図5参照)に対して、取得部16と、第2の記憶部17とが付加されている。なお、その他の構成は第1の実施形態と同様であるため、以下では、第1の実施形態と同様の構成については同一または類似の符号を付して重複した説明を省略し、本実施形態において特徴的な構成について説明する。

【0053】

取得部16は、出力部14が出力した変換後問合せ式に対するユーザの評価を反映したスコア値を取得する。具体的には、取得部16は、例えば、出力部14からクライアント

50

端末 20 に対して変換後問合せ式が出力されるのと連動して、クライアント端末 20 に、変換後問合せ式に対するスコア値を入力するための入力手段（例えば GUI 等で実現）を提供する。そして、ユーザがこの入力手段を用いて変換後問合せ式に対する評価を反映したスコア値を入力すると、入力されたスコア値が取得部 16 により取得される。取得部 16 は、このようにして取得したスコア値を変換後問合せ式と対応付けて第 2 の記憶部 17 に格納する。

【 0054 】

第 2 の記憶部 17 は、ある問合せ式に対して変換部 154' が過去に変換した変換後問合せ式を、この変換後問合せ式に対して付与されたスコア値とともに記憶するメモリである。第 2 の記憶部 17 には、出力部 14 からクライアント端末 20 に対して変換後問合せ式が出力されるたびに、その出力された変換後問合せ式に対して一定のスコア値が付与されて格納される。そして、同一の変換後問合せ式が複数回に亘って繰り返し出力された場合には、第 2 の記憶部 17 に記憶されている変換後問合せ式に対するスコア値が、その出力回数に応じて加算される。つまり、第 2 の記憶部 17 に記憶されている過去の変換後問合せ式に対するスコア値は、その変換後問合せ式の過去の出力回数が多いほど高い値となる。

10

【 0055 】

また、第 2 の記憶部 17 に記憶されている変換後問合せ式に対するスコア値は、上述したように、取得部 16 がユーザ評価を反映したスコア値を取得すると、そのユーザ評価を反映したスコア値によって書き換えられる。つまり、変換後問合せ式に対するユーザ評価がなされた場合、第 2 の記憶部 17 に記憶されている過去の変換後問合せ式に対するスコア値は、ユーザ評価が高いほど高い値となる。なお、ここではユーザ評価を重視して、第 2 の記憶部 17 に記憶されている変換後問合せ式に対するスコア値を取得部 16 が取得したスコア値によって書き換える例を示しているが、変換後問合せ式に対するユーザ評価がなされた場合に、第 2 の記憶部 17 に記憶されている変換後問合せ式に対するスコア値に、取得部 16 が取得したスコア値を加算するようによい。

20

【 0056 】

本実施形態にかかる問合せ式変換装置 10' では、問合せ式変換処理部 15' の変換部 154' は、第 1 の実施形態と同様に、元問合せ式の制約条件を緩和しながら合成構造データガイドとのマッチングにより複数の変換後問合せ式の候補を生成し、これら複数の変換後問合せ式の候補の中からスコア値が高いものを選択する。このとき、変換部 154' は、生成した変換後問合せ式の候補に対するスコア付けに、第 2 の記憶部 17 に記憶されている情報を利用する。すなわち、変換部 154' は、生成した複数の変換後問合せ式の候補に対してスコア付けを行う際に、制約条件の緩和レベルに応じたスコア値と、第 2 の記憶部 17 に同じ問合せ式が記憶されている変換後問合せ式の候補については、その問合せ式に対して付与されたスコア値とを用いてスコア付けを行う。

30

【 0057 】

図 11 は、本実施形態にかかる問合せ式変換装置 10' の変換部 154' による処理の具体的な一例を説明する図である。この図 11 に示す例は、図 7 に示した例と同様の条件で元問合せ式を変換する場合を示している。つまり、入力部 13 が入力した元問合せ式が / A / B / C であり、制約条件の緩和の例として、(1) / を / / に緩和する場合、(2) 未知要素を新規追加する場合、(3) 同一構造内の要素を OR で繋ぐ場合、を想定し、(1) の制約緩和コストを 3、(2) の制約緩和コストを 7、(3) の制約緩和コストを 5 としている。また、元問合せ式の初期スコアを 100 点としている。

40

【 0058 】

図 11 の例の場合、元問合せ式に対して (1) の制約緩和ルールを適用した / A / / B / C は、過去に変換後問合せ式として選択された問合せ式であり、スコア値 + 5 が付与されて第 2 の記憶部 17 に記憶されているものとする。また、元問合せ式に対して (2) の制約緩和ルールを適用した / A / G / B / C は、過去に変換後問合せ式として複数回選択された問合せ式であり、スコア値 + 10 が付与されて第 2 の記憶部 17 に記憶されている

50

ものとする。また、元問合せ式に対して(3)の制約緩和ルールを適用した/A/(B|C|D)/Cは、過去に変換後問合せ式として選択されたことがなく、第2の記憶部17には記憶されていない問合せ式であるとする。

【0059】

この場合、元問合せ式に対して(1)の制約緩和ルールを適用した/A//B/Cのスコア値は、 $100 - 3 + 5 = 102$ となる。また、元問合せ式に対して(2)の制約緩和ルールを適用した/A/G/B/Cのスコア値は、 $100 - 7 + 10 = 103$ となる。また、元問合せ式に対して(3)の制約緩和ルールを適用した/A/(B|C|D)/Cのスコア値は、 $100 - 5 + 0 = 95$ となる。したがって、以上の3つの変換後問合せ式の候補の中では、元問合せ式に対して(2)の制約緩和ルールを適用した/A/G/B/Cのスコア値が最も高いので、変換部154'は、これらの候補の中から変換後問合せ式として/A/G/B/Cを選択し、入力部13が入力した元問合せ式/A/B/Cを/A/G/B/Cに変換する。

10

【0060】

以上のように、本実施形態にかかる問合せ式変換装置10'は、過去の変換処理によって得た元問合せ式に対する変換後問合せ式を、その変換後問合せ式に対して付与されたスコア値とともに第2の記憶部17に記憶しておき、その後の変換処理において変換部154'が複数の変換後問合せ式の候補に対してスコア付けを行う際に、第2の記憶部17に記憶されたスコア値も加味してスコア付けを行うようにしている。したがって、学習効果によって処理を行うごとに問合せ式の変換精度が向上し、統合検索システムにおける検索の精度向上を実現することができる。

20

【0061】

また、本実施形態にかかる問合せ式変換装置10'は、取得部16が変換後問合せ式に対するユーザ評価を反映したスコア値を取得し、第2の記憶部17に格納するようにしているので、ユーザ評価をフィードバックして問合せ式の変換を行うことができ、対話的に問合せ式の変換精度の向上を図ることができる。

【0062】

以上、第1の実施形態にかかる問合せ式変換装置10および第2の実施形態にかかる問合せ式変換装置10'について説明したが、これら問合せ式変換装置10, 10'は、ハードウェア構成としては、例えば、CPUなどの制御装置、ROMやRAMなどの記憶装置、HDD、CDドライブ装置などの外部記憶装置を備えた、通常のコンピュータを利用したハードウェア構成を採用し、コンピュータによって実行される問合せ式変換プログラムにより、上述した第1の生成部11、入力部13、出力部14、URL照合部151、第2の生成部152、第3の生成部153、変換部154(変換部154')、取得部16などの機能構成を実現し、記憶装置あるいは外部記憶装置を用いて上述した第1の記憶部12、第2の記憶部17を実現する構成とすることができる。

30

【0063】

上述した問合せ式変換装置10, 10'で実行される問合せ式変換プログラムは、例えば、インストール可能な形式又は実行可能な形式のファイルでCD-ROM、フレキシブルディスク(FD)、CD-R、DVD(Digital Versatile Disc)などのコンピュータで読み取り可能な記録媒体に記録されて提供される。

40

【0064】

また、上述した問合せ式変換装置10, 10'で実行される問合せ式変換プログラムを、インターネットなどのネットワークに接続されたコンピュータ上に格納し、ネットワーク経由でダウンロードさせることにより提供するように構成してもよい。また、上述した問合せ式変換装置10, 10'で実行される問合せ式変換プログラムを、インターネットなどのネットワーク経由で提供または配布するように構成してもよい。さらに、上述した問合せ式変換装置10, 10'で実行される問合せ式変換プログラムを、ROMなどに予め組み込んで提供するように構成してもよい。

【0065】

50

上述した問合せ式変換装置 10, 10' で実行される問合せ式変換プログラムは、上述した各部（第 1 の生成部 11、入力部 13、出力部 14、URL 照合部 151、第 2 の生成部 152、第 3 の生成部 153、変換部 154（変換部 154'）、取得部 16）を含むモジュール構成となっており、実際のハードウェアとしては CPU（プロセッサ）が上記記憶媒体から情報提示プログラムを読み出して実行することにより上記各部が主記憶装置上にロードされ、第 1 の生成部 11、入力部 13、出力部 14、URL 照合部 151、第 2 の生成部 152、第 3 の生成部 153、変換部 154（変換部 154'）、取得部 16 が主記憶装置上に生成されるようになっている。

【0066】

以上、具体的な例を挙げながら詳細に説明したように、第 1 の実施形態にかかる問合せ式変換装置 10 および第 2 の実施形態にかかる問合せ式変換装置 10' によれば、ネットワーク上の各 DB サーバ 30 についてのスキーマに従った変換ルールを事前に定義しておくことなく、問合せの対象となる DB サーバ 30 に適合するように問合せ式を適切に変換することができ、また、スキーマが未知の DB サーバ 30 にも対応することができる。

10

【0067】

なお、本発明は、上述した実施の形態そのままに限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で構成要素を変形して具体化することができる。また、上述した実施の形態に開示されている複数の構成要素の適宜な組み合わせにより、種々の発明を形成することができる。例えば、実施の形態に示される全構成要素からいくつかの構成要素を削除してもよい。さらに、異なる実施の形態にわたる構成要素を適宜組み合わせてもよい。

20

【0068】

また、上述した実施の形態は、統合検索システムへの適用例であるが、1つのデータベースに対する検索を行う場合においても、本発明は有効に適用可能である。この場合にも、問合せ式変換装置は、問合せの対象となるデータベースについての構造データガイドを取得して、元問合せ式をこの構造データガイドに適合するように変換することにより、問合せの対象となるデータベースについてのスキーマに従った変換ルールを事前に定義しておくことなく、元問合せ式を適切に変換することができる。

【0069】

また、上述した実施の形態では、構造化データとして XML データを例示したが、本発明は、XML データに限らず、同様の構造化データを扱う場合に広く適用可能である。また、上述した実施の形態では、問合せ式変換装置をネットワーク上のサーバとして構成し、ネットワーク上のデータベースに対する検索を行う例を説明したが、1つの装置内部に分散されているデータ集合に対して検索を行う場合にも、本発明は有効に適用可能である。

30

【符号の説明】

【0070】

- 10 問合せ式変換装置
- 11 第 1 の生成部
- 12 第 1 の記憶部
- 13 入力部
- 14 出力部
- 15 問合せ式変換処理部
- 151 URL 照合部
- 152 第 2 の生成部
- 153 第 3 の生成部
- 154 変換部
- 16 取得部
- 17 第 2 の記憶部

40

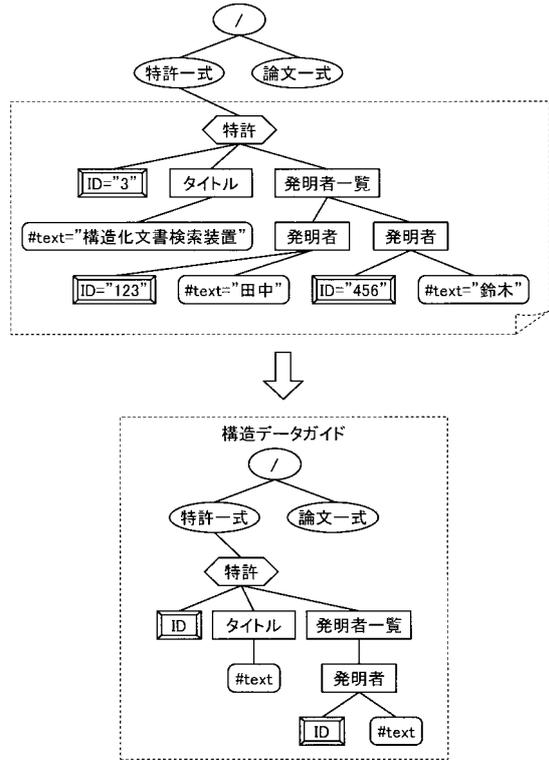
【図1】

```

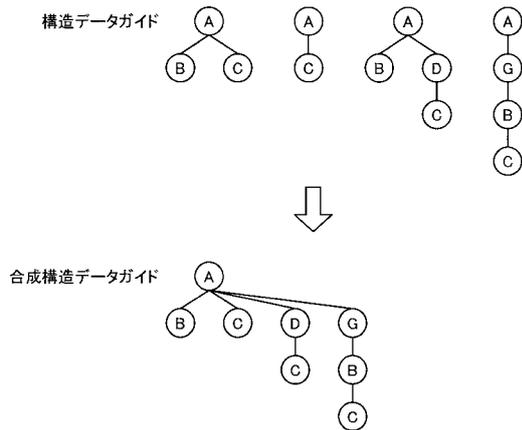
<特許ID="3">
<タイトル>構造化文書検索装置</タイトル>
<発明者リスト>
<発明者>
<姓>田中</姓>
<名>一郎</名>
</発明者>
<発明者>
<姓>鈴木</姓>
<名>次郎</名>
</発明者>
</発明者リスト>
<効果>タグ付けされた構造化文書の検索を高速化する</効果>
<キーワードリスト>
<キーワード>XML</キーワード>
<キーワード>構造化文書</キーワード>
<キーワード>データベース</キーワード>
</キーワードリスト>
</特許>

```

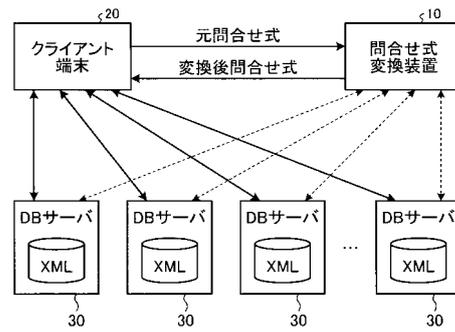
【図2】



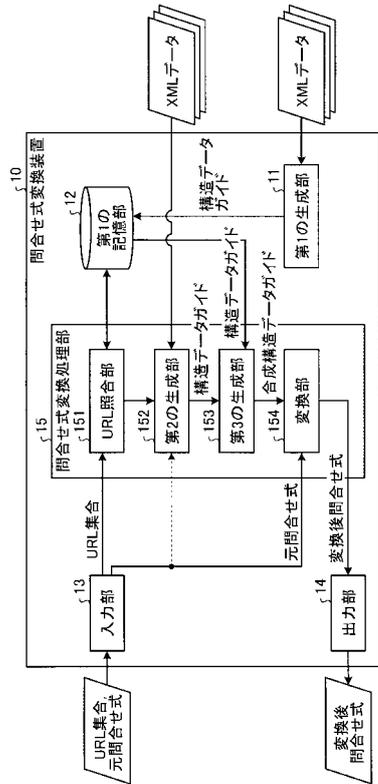
【図3】



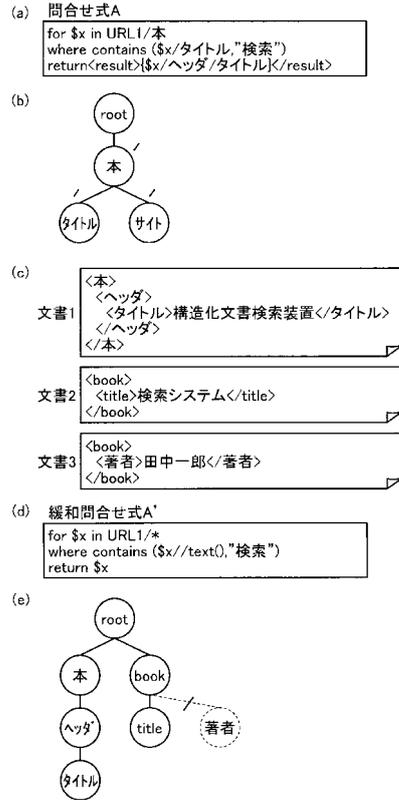
【図4】



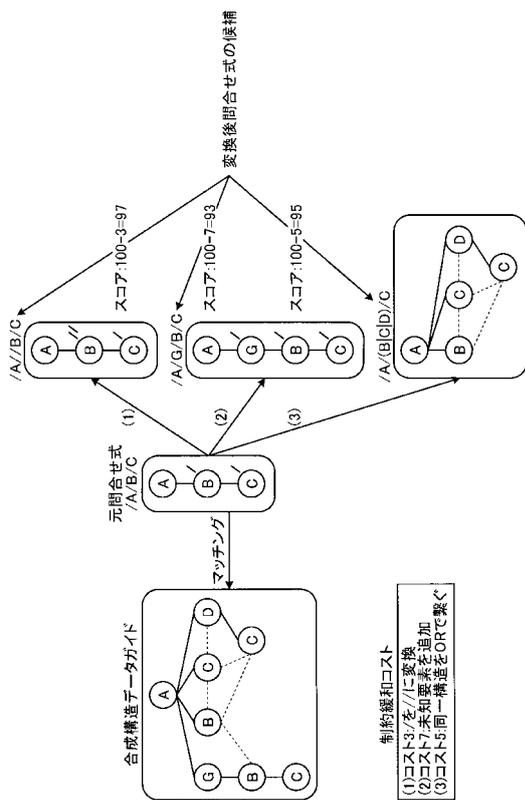
【図5】



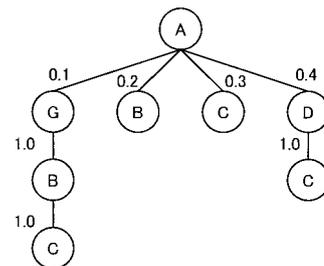
【図6】



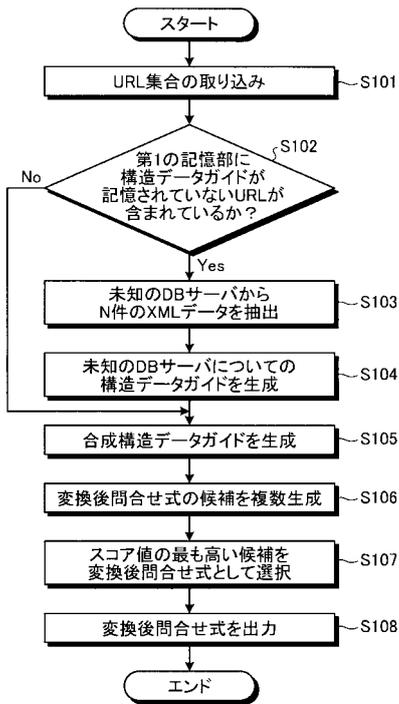
【図7】



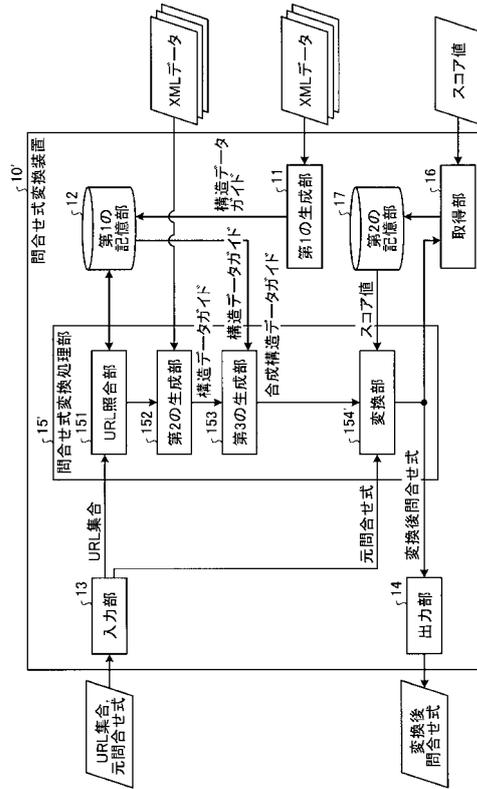
【図8】



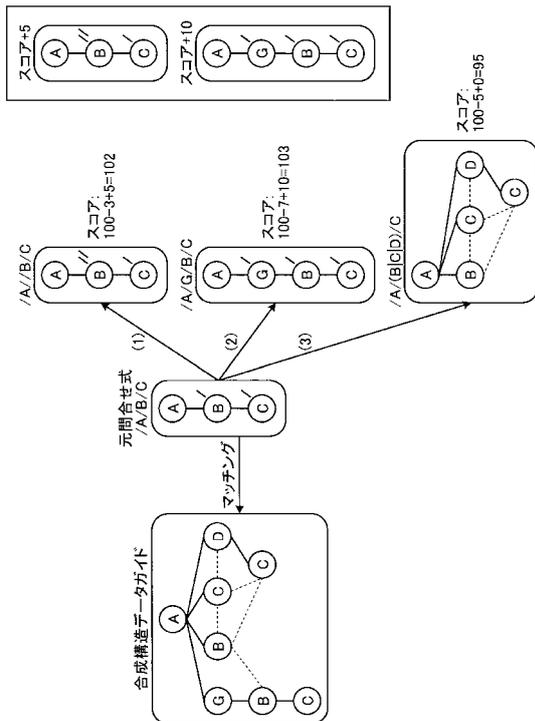
【図9】



【図10】



【図11】



フロントページの続き

(56)参考文献 特開2003-316783(JP,A)
特開2012-093829(JP,A)

(58)調査した分野(Int.Cl., DB名)
G06F 17/30