

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4547300号
(P4547300)

(45) 発行日 平成22年9月22日(2010.9.22)

(24) 登録日 平成22年7月9日(2010.7.9)

(51) Int.Cl. F I
G06F 17/30 (2006.01)
 G06F 17/30 140
 G06F 17/30 419B
 G06F 17/30 170A

請求項の数 21 (全 18 頁)

(21) 出願番号	特願2005-136706 (P2005-136706)	(73) 特許権者	000004226
(22) 出願日	平成17年5月9日(2005.5.9)		日本電信電話株式会社
(65) 公開番号	特開2006-313501 (P2006-313501A)		東京都千代田区大手町二丁目3番1号
(43) 公開日	平成18年11月16日(2006.11.16)	(74) 代理人	100083806
審査請求日	平成19年8月8日(2007.8.8)		弁理士 三好 秀和
		(72) 発明者	佐藤 宏之
			東京都千代田区大手町二丁目3番1号 日 本電信電話株式会社内
		(72) 発明者	飯塚 京士
			東京都千代田区大手町二丁目3番1号 日 本電信電話株式会社内
		(72) 発明者	向垣内 岳弥
			東京都千代田区大手町二丁目3番1号 日 本電信電話株式会社内

最終頁に続く

(54) 【発明の名称】 共通クエリグラフパターン生成装置、生成方法、生成用プログラム、およびこれらを用いた共通サブグラフ検索装置、検索方法、検索用プログラム

(57) 【特許請求の範囲】

【請求項1】

主語となるデータおよび目的語となるデータを値としたノードと、これらのノード間の関係である述語となるデータを示すプロパティとにより、グラフ構造で表現されたデータベースが格納されたグラフ構造データ格納装置に接続された共通クエリグラフパターン生成装置において、

入力されたキーワードと、選択された概念情報とを取得する検索内容取得手段と、

前記キーワードを値とするノードにつながるプロパティと前記選択された概念情報とそれにつながるプロパティとを固定し、前記キーワードに該当するノードの値を任意のデータにするとともに、前記固定された2つのプロパティの間のパスに含まれるノードの値およびプロパティを任意のデータとした検索クエリグラフパターンを生成する検索クエリグラフパターン生成手段と、

前記検索クエリグラフパターン生成手段で生成された検索クエリグラフパターンの構造にマッチする部分の情報を、前記グラフ構造データ格納装置に格納されたデータベースの中から取得する検索サブグラフ取得手段と、

前記検索サブグラフ取得手段で取得された検索サブグラフのうち、前記任意の値にしたキーワードに該当するノードにマッチした値が同一の検索サブグラフの中に複数個存在する、パスの構造およびこのパスに含まれるプロパティが一致するグラフ構造部分の情報を部分検索サブグラフとして抽出する部分検索サブグラフ抽出手段と、

前記部分検索サブグラフ抽出手段で抽出された部分検索サブグラフに含まれるノードを

任意の値にすることで、前記入力されたキーワードおよび選択された概念情報に関連し共通のグラフ構造部分を有するデータを検索するための共通クエリグラフパターンを生成する共通クエリグラフパターン生成手段と

を備えることを特徴とする共通クエリグラフパターン生成装置。

【請求項 2】

前記共通クエリグラフパターン生成装置は、等価の意味ごとに言葉が分類されたデータであるオントロジデータが格納されたオントロジデータ格納装置にさらに接続され、

前記部分検索サブグラフ抽出手段は、前記部分検索サブグラフを抽出する際に、前記オントロジデータ格納装置に格納されたオントロジデータを用いて、等価の意味として分類されている言葉に対応する複数のプロパティを、同一のプロパティとして抽出することを特徴とする請求項 1 に記載の共通クエリグラフパターン生成装置。

10

【請求項 3】

前記共通クエリグラフパターン生成手段は、生成した共通クエリグラフパターンの中の、前記任意の値にしたキーワードに該当するノードの値として、前記検索内容取得手段で取得されたキーワードを代入した共通クエリグラフパターンを生成する

ことを特徴とする請求項 1 または 2 に記載の共通クエリグラフパターン生成装置。

【請求項 4】

主語となるデータおよび目的語となるデータを値としたノードと、これらのノード間の関係である述語となるデータを示すプロパティとにより、グラフ構造で表現されたデータベースが格納されたグラフ構造データ格納装置に接続された共通クエリグラフパターン生成装置が、

20

入力されたキーワードと、選択された概念情報とを取得する検索内容取得ステップと、

前記キーワードを値とするノードにつながるプロパティと前記選択された概念情報とそれにつながるプロパティとを固定し、前記キーワードに該当するノードを任意の値にするとともに、前記固定された 2 つのプロパティの間のパスに含まれるノードの値およびプロパティを任意のデータとした検索クエリグラフパターンを生成する検索クエリグラフパターン生成ステップと、

前記検索クエリグラフパターン生成ステップで生成された検索クエリグラフパターンの構造にマッチする部分の情報を、前記グラフ構造データ格納装置に格納されたデータベースの中から取得する検索サブグラフ取得ステップと、

30

前記検索サブグラフ取得ステップで取得された検索サブグラフのうち、前記任意の値にしたキーワードに該当するノードにマッチした値が同一の検索サブグラフの中に複数個存在する、パスの構造およびこのパスに含まれるプロパティが一致するグラフ構造部分の情報を部分検索サブグラフとして抽出する部分検索サブグラフ抽出ステップと、

前記部分検索サブグラフ抽出ステップで抽出された部分検索サブグラフに含まれるノードを任意の値にすることで、前記入力されたキーワードおよび選択された概念情報に関連し共通のグラフ構造部分を有するデータを検索するための共通クエリグラフパターンを生成する共通クエリグラフパターン生成ステップと

を有することを特徴とする共通クエリグラフパターン生成方法。

【請求項 5】

40

前記共通クエリグラフパターン生成装置は、等価の意味ごとに言葉が分類されたデータであるオントロジデータが格納されたオントロジデータ格納装置にさらに接続され、

前記部分検索サブグラフ抽出ステップでは、前記部分検索サブグラフを抽出する際に、前記オントロジデータ格納装置に格納されたオントロジデータを用いて、等価の意味として分類されている言葉に対応する複数のプロパティを、同一のプロパティとして抽出することを特徴とする請求項 4 に記載の共通クエリグラフパターン生成方法。

【請求項 6】

前記共通クエリグラフパターン生成ステップでは、生成した共通クエリグラフパターンの中の、前記任意の値にしたキーワードに該当するノードの値として、前記検索内容取得ステップで取得されたキーワードを代入した共通クエリグラフパターンを生成する

50

ことを特徴とする請求項4または5に記載の共通クエリグラフパターン生成方法。

【請求項7】

主語となるデータおよび目的語となるデータを値としたノードと、これらのノード間の関係である述語となるデータを示すプロパティとにより、グラフ構造で表現されたデータベースが格納されたグラフ構造データ格納装置に接続された共通クエリグラフパターン生成装置に、

入力されたキーワードと、選択された概念情報とを取得する検索内容取得機能と、

前記キーワードを値とするノードにつながるプロパティと前記選択された概念情報とそれにつながるプロパティとを固定し、前記キーワードに該当するノードを任意の値にするとともに、前記固定された2つのプロパティの間のパスに含まれるノードの値およびプロ

10

パティを任意のデータとした検索クエリグラフパターンを生成する検索クエリグラフパターン生成機能と、

前記検索クエリグラフパターン生成機能により生成された検索クエリグラフパターンの構造にマッチする部分の情報を、前記グラフ構造データ格納装置に格納されたデータベースの中から取得する検索サブグラフ取得機能と、

前記検索サブグラフ取得機能により取得された検索サブグラフのうち、前記任意の値にしたキーワードに該当するノードにマッチした値が同一の検索サブグラフの中に複数個存在する、パスの構造およびこのパスに含まれるプロパティが一致するグラフ構造部分の情報を部分検索サブグラフとして抽出する部分検索サブグラフ抽出機能と、

20

前記部分検索サブグラフ抽出機能により抽出された部分検索サブグラフに含まれるノードを任意の値にすることで、前記入力されたキーワードおよび選択された概念情報に関連し共通のグラフ構造部分を有するデータを検索するための共通クエリグラフパターンを生成する共通クエリグラフパターン生成機能と

【請求項8】

前記共通クエリグラフパターン生成装置は、等価の意味ごとに言葉が分類されたデータであるオントロジデータが格納されたオントロジデータ格納装置にさらに接続され、

前記部分検索サブグラフ抽出機能により、前記部分検索サブグラフを抽出する際に、前記オントロジデータ格納装置に格納されたオントロジデータを用いて、等価の意味として分類されている言葉に対応する複数のプロパティを、同一のプロパティとして抽出する

30

【請求項9】

前記共通クエリグラフパターン生成機能により、生成した共通クエリグラフパターンの中の、前記任意の値にしたキーワードに該当するノードの値として、前記検索内容取得機能により取得されたキーワードを代入した共通クエリグラフパターンを生成する

ことを特徴とする請求項7または8に記載の共通クエリグラフパターン生成用プログラム

【請求項10】

主語となるデータおよび目的語となるデータを値としたノードと、これらのノード間の関係である述語となるデータを示すプロパティとにより、グラフ構造で表現されたデータベースが格納されたグラフ構造データ格納装置に接続された共通サブグラフ検索装置において、

40

入力されたキーワードと、選択された概念情報とを取得する検索内容取得手段と、

前記キーワードを値とするノードにつながるプロパティと前記選択された概念情報とそれにつながるプロパティとを固定し、前記キーワードに該当するノードを任意の値にするとともに、前記固定された2つのプロパティの間のパスに含まれるノードの値およびプロパティを任意のデータとした検索クエリグラフパターンを生成する検索クエリグラフパターン生成手段と、

前記検索クエリグラフパターン生成手段で生成された検索クエリグラフパターンの構造にマッチする部分の情報を、前記グラフ構造データ格納装置に格納されたデータベースの

50

中から取得する検索サブグラフ取得手段と、

前記検索サブグラフ取得手段で取得された検索サブグラフのうち、前記任意の値にしたキーワードに該当するノードにマッチした値が同一の検索サブグラフの中に複数個存在する、パスの構造およびこのパスに含まれるプロパティが一致するグラフ構造部分の情報を部分検索サブグラフとして抽出する部分検索サブグラフ抽出手段と、

前記部分検索サブグラフ抽出手段で抽出された部分検索サブグラフに含まれるノードを任意の値にすることで、前記入力されたキーワードおよび選択された概念情報に関連し共通のグラフ構造部分を有するデータを検索するための共通クエリグラフパターンを生成する共通クエリグラフパターン生成手段と、

前記共通クエリグラフパターン生成手段で生成された共通クエリグラフパターンの構造にマッチする部分の情報を、前記グラフ構造データ格納装置に格納されたデータベースの中から共通サブグラフとして取得する共通サブグラフ取得手段と、

を備えることを特徴とする共通サブグラフ検索装置。

【請求項 1 1】

前記共通サブグラフ検索装置は、等価の意味ごとに言葉が分類されたデータであるオントロジデータが格納されたオントロジデータ格納装置にさらに接続され、

前記部分検索サブグラフ抽出手段は、前記部分検索サブグラフを抽出する際に、前記オントロジデータ格納装置に格納されたオントロジデータを用いて、等価の意味として分類されている言葉に対応する複数のプロパティを、同一のプロパティとして抽出することを特徴とする請求項 1 0 に記載の共通サブグラフ検索装置。

【請求項 1 2】

前記共通クエリグラフパターン生成手段は、生成した共通クエリグラフパターンの中の、前記任意の値にしたキーワードに該当するノードの値として、前記検索内容取得手段で取得されたキーワードを代入した共通クエリグラフパターンを生成する

ことを特徴とする請求項 1 0 または 1 1 に記載の共通サブグラフ検索装置。

【請求項 1 3】

前記共通サブグラフ取得手段で取得された共通サブグラフを、前記共通クエリグラフパターンの構造にマッチした情報ごとに、各ノードの値および各プロパティで示されるデータを表形式で表示させるためのデータに変換する表示データ変換手段をさらに備える

ことを特徴とする請求項 1 0 ~ 1 2 何れか 1 項に記載の共通サブグラフ検索装置。

【請求項 1 4】

主語となるデータおよび目的語となるデータを値としたノードと、これらのノード間の関係である述語となるデータを示すプロパティとにより、グラフ構造で表現されたデータベースが格納されたグラフ構造データ格納装置に接続された共通サブグラフ検索装置が、

入力されたキーワードと、選択された概念情報とを取得する検索内容取得ステップと、

前記キーワードを値とするノードにつながるプロパティと前記選択された概念情報とそれにつながるプロパティとを固定し、前記キーワードに該当するノードを任意の値にする

とともに、前記固定された 2 つのプロパティの間のパスに含まれるノードの値およびプロパティを任意のデータとした検索クエリグラフパターンを生成する検索クエリグラフパターン生成ステップと、

前記検索クエリグラフパターン生成ステップで生成された検索クエリグラフパターンの構造にマッチする部分の情報を、前記グラフ構造データ格納装置に格納されたデータベースの中から取得する検索サブグラフ取得ステップと、

前記検索サブグラフ取得ステップにおいて取得された検索サブグラフのうち、前記任意の値にしたキーワードに該当するノードにマッチした値が同一の検索サブグラフの中に複数個存在する、パスの構造およびこのパスに含まれるプロパティが一致するグラフ構造部分の情報を部分検索サブグラフとして抽出する部分検索サブグラフ抽出ステップと、

前記部分検索サブグラフ抽出ステップにおいて抽出された部分検索サブグラフに含まれるノードを任意の値にすることで、前記入力されたキーワードおよび選択された概念情報に関連し共通のグラフ構造部分を有するデータを検索するための共通クエリグラフパター

10

20

30

40

50

ンを生成する共通クエリグラフパターン生成ステップと、

前記共通クエリグラフパターン生成ステップにおいて生成された共通クエリグラフパターンの構造にマッチする部分の情報を、前記グラフ構造データ格納装置に格納されたデータベースの中から共通サブグラフとして取得する共通サブグラフ取得ステップと、
を有することを特徴とする共通サブグラフ検索方法。

【請求項 15】

前記共通サブグラフ検索装置は、等価の意味ごとに言葉が分類されたデータであるオン
トログジデータが格納されたオントログジデータ格納装置にさらに接続され、

前記部分検索サブグラフ抽出ステップでは、前記部分検索サブグラフを抽出する際に、
前記オントログジデータ格納装置に格納されたオントログジデータを用いて、等価の意味として分類されている言葉に対応する複数のプロパティを、同一のプロパティとして抽出することを特徴とする請求項 14 に記載の共通サブグラフ検索方法。

10

【請求項 16】

前記共通クエリグラフパターン生成ステップでは、生成した共通クエリグラフパターン
の中の、前記任意の値にしたキーワードに該当するノードの値として、前記検索内容取得
ステップで取得されたキーワードを代入した共通クエリグラフパターンを生成する
ことを特徴とする請求項 14 または 15 に記載の共通サブグラフ検索方法。

【請求項 17】

前記共通サブグラフ取得ステップで取得された共通サブグラフを、前記共通クエリグラ
フパターンの構造にマッチした情報ごとに、各ノードの値および各プロパティで示される
データを表形式で表示させるためのデータに変換する表示データ変換ステップをさらに有
する

20

ことを特徴とする請求項 14 ~ 16 何れか 1 項に記載の共通サブグラフ検索方法。

【請求項 18】

主語となるデータおよび目的語となるデータを値としたノードと、これらのノード間の
関係である述語となるデータを示すプロパティとにより、グラフ構造で表現されたデータ
ベースが格納されたグラフ構造データ格納装置に接続された共通サブグラフ検索装置に、

入力されたキーワードと、選択された概念情報とを取得する検索内容取得機能と、

前記キーワードを値とするノードにつながるプロパティと前記選択された概念情報とそ
れにつながるプロパティとを固定し、前記キーワードに該当するノードを任意の値にする
とともに、前記固定された 2 つのプロパティの間のパスに含まれるノードの値およびプロ
パティを任意のデータとした検索クエリグラフパターンを生成する検索クエリグラフパ
ターン生成機能と、

30

前記検索クエリグラフパターン生成機能により生成された検索クエリグラフパターンの
構造にマッチする部分の情報を、前記グラフ構造データ格納装置に格納されたデータベ
ースの中から取得する検索サブグラフ取得機能と、

前記検索サブグラフ取得機能により取得された検索サブグラフのうち、前記任意の値に
したキーワードに該当するノードにマッチした値が同一の検索サブグラフの中に複数個存
在する、パスの構造およびこのパスに含まれるプロパティが一致するグラフ構造部分の情
報を部分検索サブグラフとして抽出する部分検索サブグラフ抽出機能と、

40

前記部分検索サブグラフ抽出機能により抽出された部分検索サブグラフに含まれるノ
ードを任意の値にすることで、前記入力されたキーワードおよび選択された概念情報に関連
し共通のグラフ構造部分を有するデータを検索するための共通クエリグラフパターンを生
成する共通クエリグラフパターン生成機能と、

前記共通クエリグラフパターン生成機能により生成された共通クエリグラフパターンの
構造にマッチする部分の情報を、前記グラフ構造データ格納装置に格納されたデータベ
ースの中から共通サブグラフとして取得する共通サブグラフ取得機能と、

を実行させることを特徴とする共通サブグラフ検索用プログラム。

【請求項 19】

前記共通サブグラフ検索装置は、等価の意味ごとに言葉が分類されたデータであるオン

50

トロジデータが格納されたオントロジデータ格納装置にさらに接続され、

前記部分検索サブグラフ抽出機能により、前記部分検索サブグラフを抽出する際に、前記オントロジデータ格納装置に格納されたオントロジデータを用いて、等価の意味として分類されている言葉に対応する複数のプロパティを、同一のプロパティとして抽出することを特徴とする請求項 18 に記載の共通サブグラフ検索用プログラム。

【請求項 20】

前記共通クエリグラフパターン生成機能では、生成した共通クエリグラフパターンの中の、前記任意の値にしたキーワードに該当するノードの値として、前記検索内容取得機能により取得されたキーワードを代入した共通クエリグラフパターンを生成することを特徴とする請求項 18 または 19 に記載の共通サブグラフ検索用プログラム。

10

【請求項 21】

前記共通サブグラフ取得機能により取得された共通サブグラフを、前記共通クエリグラフパターンの構造にマッチした情報ごとに、各ノードの値および各プロパティで示されるデータを表形式で表示させるためのデータに変換する表示データ変換機能をさらに有することを特徴とする請求項 18 ~ 20 何れか 1 項に記載の共通サブグラフ検索用プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、膨大なデータの中からユーザが必要とする情報を検索するために用いる共通クエリグラフパターン生成装置、生成方法、生成用プログラム、およびこれらを用いた共通サブグラフ検索装置、検索方法、検索用プログラムに関する。

20

【背景技術】

【0002】

近年、コンピュータネットワーク上には大量のデータが存在しており、ユーザはそこから所望のデータを検索することが困難になってきている。そのため、大量のデータから所望のデータを簡易に入手するための精度の高い検索技術が注目を集めている。

【0003】

このような精度の高い検索を行うために用いられる技術にセマンティックWebがある。このセマンティックWebは、Webサイト上にメタデータとしてセマンティクス（意味情報）を付与し、人の手を使わずにデータを機械的に処理できる空間を創出するものである。

30

【0004】

このセマンティックWebを利用した技術が非特許文献 1 に記載されている。この文献には、セマンティックWebを利用し、メタデータの記述方法を定めたResource Description Framework (RDF) により表現されたグラフ構造のデータから、グラフパターンをクエリとしてマッチングさせて必要なデータを検索する技術が記載されている。またこの文献では、マッチした部分をサブグラフとして取得するための仕様が規定されている。

【非特許文献 1】 SPARQL Query Language for RDF URL : <http://www.w3.org/TR/RDF-sparql-query/>

【非特許文献 2】 gSpan: Graph-Based Substructure Pattern Mining, Proceedings of the 2002 IEEE International Conference on Data Mining(ICDM '02), IEEE Computer Society, Washington, DC, USA

40

【発明の開示】

【発明が解決しようとする課題】

【0005】

しかし、この非特許文献 1 に記載の技術では、グラフ構造を持つ大量のデータの中から所望のサブグラフを抽出する際、予め抽出したいサブグラフの構造が分かっていたらならなかった。これは、ユーザの検索履歴データやユーザプロフィールデータの情報があってもこれらの情報からクエリグラフパターンを作成する方法については明示されていないため、ユーザ自身がそのグラフパターンをクエリとして入力しなければならないからで

50

ある。そのため、ユーザによって異なる興味や関心に応じてクエリグラフパターンを変更する場合も、ユーザ自身が変更するグラフパターンを直接指定しなければならず、手間がかかるという問題があった。

【0006】

また、非特許文献2には、グラフ集合から頻出部分のサブグラフを検出するアルゴリズムが記載されている。しかしこの技術は、グラフ構造を持った膨大な情報の中から頻出のサブグラフ構造を検出することはできるが、その際ユーザの興味対象に関連づけられた構造に基づいて検出することはできない。

【0007】

本発明は、上記事情に鑑みてなされたものであり、グラフ構造を持つ大量のデータの中からクエリグラフパターンが一致する情報を検索する際、ユーザにより入力された情報または予め設定された情報に関連し且つ共通の意味合いの構造を持つ情報を取得することができるクエリグラフパターンを自動的に生成できる共通クエリグラフパターン生成装置、生成方法、生成用プログラム、およびこれらを用いた共通サブグラフ検索装置、検索方法、検索性プログラムを提供することを目的とする。

【課題を解決するための手段】

【0008】

上記目的を達成するための本発明は、主語となるデータおよび目的語となるデータを値としたノードと、これらのノード間の関係である述語となるデータを示すプロパティとにより、グラフ構造で表現されたデータベースが格納されたグラフ構造データ格納装置に接続された共通クエリグラフパターン生成装置において、入力されたキーワードと、選択された概念情報とを取得する検索内容取得手段と、前記グラフ構造データ格納装置に格納されたデータベース内から、前記キーワードを値とするノードにつながるプロパティと前記選択された概念情報とそれにつながるプロパティとを固定し、前記キーワードに該当するノードを任意の値にするとともに、前記固定された2つのプロパティの間のパスに含まれるノードの値およびプロパティを任意のデータとした検索クエリグラフパターンを生成する検索クエリグラフパターン生成手段と、前記検索クエリグラフパターン生成手段で生成された検索クエリグラフパターンの構造にマッチする部分の情報を、前記グラフ構造データ格納装置に格納されたデータベースの中から取得する検索サブグラフ取得手段と、前記検索サブグラフ取得手段で取得された検索サブグラフのうち、前記任意の値にしたキーワードに該当するノードにマッチした値が同一の検索サブグラフの中に複数個存在する、パスの構造およびこのパスに含まれるプロパティが一致するグラフ構造部分の情報を部分検索サブグラフとして抽出する部分検索サブグラフ抽出手段と、前記部分検索サブグラフ抽出手段で抽出された部分検索サブグラフに含まれるノードを任意の値にすることで、前記入力されたキーワードおよび選択された概念情報に関連し共通のグラフ構造部分を有するデータを検索するための共通クエリグラフパターンを生成する共通クエリグラフパターン生成手段とを備えることを特徴とする。

【0009】

また、前記共通クエリグラフパターン生成装置は、等価の意味ごとに言葉が分類されたデータであるオントロジデータが格納されたオントロジデータ格納装置にさらに接続され、前記部分検索サブグラフ抽出手段は、前記部分検索サブグラフを抽出する際に、前記オントロジデータ格納装置に格納されたオントロジデータを用いて、等価の意味として分類されている言葉に対応する複数のプロパティを、同一のプロパティとして抽出するようにしてもよい。

【0010】

また、前記共通クエリグラフパターン生成手段は、生成した共通クエリグラフパターンの中の、前記任意の値にしたキーワードに該当するノードの値として、前記検索内容取得手段で取得されたキーワードを代入した共通クエリグラフパターンを生成するようにしてもよい。

【0011】

10

20

30

40

50

また、この共通クエリグラフパターンにより前記グラフ構造データ格納装置に格納されたデータベースの中から取得された、検索内容の情報として入力されたキーワードおよび予め設定された中から選択された属性情報に関連する情報に共通のデータである共通サブグラフを、前記共通クエリグラフパターンの構造にマッチした情報ごとに、各ノードの値および各プロパティで示されるデータを表形式で表示させるためのデータに変換する表示データ変換手段をさらに備えるようにしてもよい。

【発明の効果】

【0012】

本発明によれば、グラフ構造を持つ大量のデータの中からクエリグラフパターンが一致する情報を検索する際に、ユーザによりキーワードまたは概念などの情報が入力されることにより、これらの情報に関連し且つ共通の意味合いの構造を持つ情報を取得することができるクエリグラフパターンを自動的に生成することができるため、所望の情報を効率よく取得することができる。

10

【発明を実施するための最良の形態】

【0013】

本発明の共通クエリグラフパターン生成装置の一実施形態を、図1～図19を用いて説明する。

【0014】

《一実施形態による検索システム1の構成》

図1は、本発明の一実施形態による共通クエリグラフパターン生成装置としての共通クエリグラフ生成部13を有する検索装置10を用いた検索システム1の構成を示す全体図である。

20

【0015】

本実施形態における検索システム1は、検索装置10と、ユーザ端末20と、グラフ構造データベース30と、オントロジデータベース40とを有する。

【0016】

検索装置10は、ユーザ端末20、グラフ構造データベース30、およびオントロジデータベース40にアクセス可能な状態で接続されている。

【0017】

また、検索装置10は、ユーザに検索内容の情報を入力させるための入力用GUIを生成する入力用GUI生成部11と、この入力用GUIに入力された情報に基づいて共通クエリグラフパターンを生成してこの共通クエリグラフパターンにマッチする情報をグラフ構造データベースから取得する共通クエリグラフパターン生成部12と、生成された共通クエリグラフパターンを蓄積する共通クエリグラフパターン蓄積部13と、ユーザに検索結果を提示するための出力用GUIを生成する出力用GUI生成部14とを有する。

30

【0018】

グラフ構造データベース30は、図2に示すようなグラフ構造を持つデータが格納されている。本実施形態におけるグラフ構造データはラベル付き有向グラフデータであり、データのアーキ(ノード間を結ぶ矢印)をRDFの仕様に基づいてプロパティと称する。

【0019】

図3は、これらのグラフ構造データがどのように既存のリレーショナルデータベースなどで管理されているデータから生成され、グラフ化されているかを示す説明図である。図3は図2のデータの一部を示したものであり、ドキュメント「T-041」とそのドキュメントの著者「山田太郎」と、そのドキュメントの題名「セマンティックWeb入門」と、そのドキュメントのキーワード「セマンティックWeb」とが与えられているデータを、XML形式およびグラフ形式を用いてRDFで表現したものである。

40

【0020】

オントロジデータベース40は、図4に示すようなWeb Ontology Language(OWL)を用いて表現される、言葉の意味や分類体系を定義するためのオントロジデータが格納されている。図4では、プロパティ「rm:キーワード」とプロパティ「ms:用語」とは等価であり、

50

プロパティ「rm:著者」とプロパティ「ms:著者」とは等価であることが示されている。

【0021】

《一実施形態による検索システム1の動作》

次に、上記のように構成された検索システム1の動作を、図5のシーケンス図を用いて説明する。

【0022】

まず、入力用GUI生成部11において入力用GUIが生成される(S1)。本実施形態において生成される入力用GUIには、図6に示すように、キーワードを入力するスペースと、概念(クラス)を選択する欄が設けられている。

【0023】

そして、生成された入力用GUIがユーザ端末20に表示され、ユーザによりキーワードの入力および概念が選択される(S2)。本実施形態においては、ユーザによりキーワードとして「セマンティックWeb」が入力され、概念(クラス)として「人」が選択されている。入力されたキーワードおよび選択された概念のデータは共通クエリグラフパターン生成部12に出力される。

【0024】

共通クエリグラフパターン生成部12では、取得したキーワードを基にグラフ構造データベース30にアクセスが行われる(S3)。

【0025】

そして、グラフ構造データベース30において、このキーワード「セマンティックWeb」につながるプロパティが抽出される(S4)。抽出されたプロパティの弧が太線で示され、プロパティ名に下線が引かれたグラフ構造データの例を図7に示す。

【0026】

さらにグラフ構造データベース30において、これらの太線で示されたプロパティ「rm:キーワード」または「ms:用語」を基に、このプロパティ「rm:キーワード」または「ms:用語」と、ユーザにより選択されたクラス「Person:人」のインスタンス「Person:山田太郎」、「Person:田中一郎」、または「Person:鈴木花子」との間にあるパスをグラフ構造データベース30から抽出することが可能な検索クエリグラフパターンが全て生成される(S5)。

【0027】

図8は、生成された検索クエリグラフパターンを示す説明図である。本実施形態においては図8に示すように、ステップS4で抽出されたプロパティ「rm:キーワード」または「ms:用語」と、ユーザにより選択されたクラス「Person:人」と、このクラス「Person:人」のインスタンス「Person:山田太郎」、「Person:田中一郎」、または「Person:鈴木花子」からクラス「Person:人」につながるプロパティ「rdf:type」とが固定された2種類の検索クエリグラフパターンが生成されている。図8において、名称の先頭に「?」マークが記載されているノードは、任意の値が入る可能性がある変数であることを示している。

【0028】

次に、生成された2種類の検索クエリグラフパターンを基に、共通クエリグラフパターン生成部12からグラフ構造データベース30にアクセスされる(S6)。

【0029】

そして、グラフ構造データベース30においてこの検索クエリグラフパターンにマッチする検索サブグラフ(変数ノード「?keyword」と変数ノード「?Person」との間にパスが存在するサブグラフ)が抽出される(S7)。

【0030】

図9は、グラフ構造データベース30から抽出された検索サブグラフの一例である第1検索サブグラフを示す説明図である。この第1検索サブグラフでは、変数ノード「?keyword」としてノード「XML」がマッチした検索サブグラフを示す。このとき、変数ノード「?Person」の値が同一のデータについては同一ノードを1つのみ表示するようにする。

10

20

30

40

50

【 0 0 3 1 】

このように、変数ノード「?keyword」がノード「XML」であり同一の変数ノード「?Person」ごとに示された第1検索サブグラフは、図9に示すとおり(1)変数ノード「?Person」がノード「Person:山田太郎」の場合、(2)変数ノード「?Person」がノード「Person:鈴木花子」の場合、(3)変数ノード「?Person」がノード「Person:中村二郎」の場合の3種類である。

【 0 0 3 2 】

次に、これら3種類の第1検索サブグラフの中から、2点(ノード「XML」および変数ノード「?Person」)のノード間が2個以上の複数のパスで構成され、この複数のパスで構成されたグラフの構造およびこのパスのプロパティが一致し、且つこの一致した複数のパスで構成されたグラフの構造およびこのパスのプロパティが複数の検索結果の中に存在する部分第1検索サブグラフが抽出される(S8)。

10

【 0 0 3 3 】

図10は、第1検索サブグラフから抽出された部分第1検索サブグラフのプロパティを太線で示した状態を示す説明図である。本実施形態においては、ノード「XML」に2本のプロパティ「rm:キーワード」がつながり、変数ノード「?Person」にプロパティ「pj:担当者」とプロパティ「rm:著者」とがつながっている構造の部分検索サブグラフが抽出されている。

【 0 0 3 4 】

そして、この部分第1検索サブグラフにおいて、共通した構造およびプロパティ以外のノード「XML」、「Aプロジェクト」、「rm:T-041」、「Bプロジェクト」、「rm:T-057」、「Person:山田太郎」、および「Person:鈴木花子」を変数ノードとした第1共通クエリグラフパターンが共通クエリグラフパターン生成部12において生成される(S9)。

20

【 0 0 3 5 】

生成された第1共通クエリグラフパターンは、共通クエリグラフパターン生成部12から共通クエリグラフパターン蓄積部13に送出され蓄積される(S10)。

【 0 0 3 6 】

図11は、生成された第1共通クエリグラフパターンを示す説明図である。本実施形態においては図11に示すように、ステップS8で抽出されたプロパティ「rm:キーワード」、「pj:担当者」、および「tm:著者」が固定された第1共通クエリグラフパターンが生成されている。

30

【 0 0 3 7 】

一方、ステップS7において、変数ノード「?keyword」としてノード「セマンティックWeb」がマッチした第2検索サブグラフについて説明する。

【 0 0 3 8 】

図12は、変数ノード「?keyword」としてノード「セマンティックWeb」がマッチした第2検索サブグラフを示す説明図である。ここでも、変数ノード「?Person」の値が同一のデータについては同一ノードを1つのみ表示するようにする。

【 0 0 3 9 】

このように、変数ノード「?keyword」がノード「セマンティックWeb」であり同一の変数ノード「?Person」ごとに示された第2検索サブグラフは、図12に示すとおり(1)変数ノード「?Person」がノード「Person:山田太郎」の場合、(2)変数ノード「?Person」がノード「Person:田中一郎」の場合、(3)変数ノード「?Person」がノード「Person:鈴木花子」の場合の3種類である。

40

【 0 0 4 0 】

次に、この第2検索サブグラフに含まれるプロパティを基に、共通クエリグラフパターン生成部12からオントロジデータベース40にアクセスされる(S11)。

【 0 0 4 1 】

そして、オントロジデータベース40において、これらの第2検索サブグラフに含まれるプロパティの中に同じ意味に分類されているものがあるかどうか検索される。ここでは

50

、第2検索サブグラフのプロパティに含まれ、同じ意味に分類されている「キーワード」と「用語」とが抽出され、共通クエリグラフパターン生成部12に送出される(S12)。

【0042】

次に、3種類の第2検索サブグラフの中から、2点(ノード「セマンティックWeb」および変数ノード「?Person」)のノード間が2個以上の複数のパスで構成され、この複数のパスで構成されたグラフの構造およびこのパスのプロパティが一致し、且つこの一致した複数のパスで構成されたグラフの構造およびこのパスのプロパティが複数の検索結果の中に存在する部分第2検索サブグラフが抽出される。このとき、ステップS14で抽出された単語は同一とみなされ、部分第2検索サブグラフが抽出される(S8)。

10

【0043】

図13は、第2検索サブグラフから抽出された部分第2検索サブグラフのプロパティを太線で示した状態を示す説明図である。本実施形態においては、ノード「セマンティックWeb」に2本のプロパティ「rm:キーワード」または「ms:用語」が繋がり、変数ノード「?Person」にプロパティ「rm:著者」またはプロパティ「ms:著者」とが繋がっている構造の部分第2検索サブグラフが抽出されている。

【0044】

そして、この部分第2検索サブグラフにおいて、共通した構造およびプロパティ以外のノード「セマンティックWeb」、「rm:T-041」、「ms:PF1-14」、「rm:T-053」、「rm:T-048」、「Person:山田太郎」、および「Person:田中一郎」を変数ノードとした第2共通クエリグラフパターンが共通クエリグラフパターン生成部12において生成される(S9)。

20

【0045】

生成された第2共通クエリグラフパターンも、共通クエリグラフパターン生成部12から共通クエリグラフパターン蓄積部13に送出され蓄積される(S10)。

【0046】

図14は、生成された第2共通クエリグラフパターンを示す説明図である。本実施形態においては図14に示すように、ステップS8で抽出されたプロパティ「rm:キーワード」、「rm:キーワード」もしくは「ms:用語」、「rm:著者」、「rm:著者」もしくは「ms:著者」が固定された第2共通クエリグラフパターンが生成されている。

【0047】

30

次に、図15に示すように、ステップS9で生成された第1クエリグラフパターンの変数ノード「?keyword」にユーザにより入力されたキーワード「セマンティックWeb」を代入し、この第1クエリグラフパターンを基に、共通クエリグラフパターン生成部12からグラフ構造データベース30にアクセスされる(S13)。

【0048】

そして、グラフ構造データベース30においてこのキーワード「セマンティックWeb」が代入された第1クエリグラフパターンにマッチする第1共通サブグラフが抽出される(S14)。

【0049】

図16は、グラフ構造データベース30から抽出された第1共通サブグラフを示す説明図である。本実施形態においては抽出された第1共通サブグラフは1件のみであり、これはユーザにより入力されたキーワード「セマンティックWeb」以外の技術(ここでは「XML」)においてみられたクラス「人」とクラス「技術」との間に共通して存在する強い関連性のある情報が、キーワード「セマンティックWeb」に関しては1件しかなかったことを示している。

40

【0050】

次に、図17に示すように、ステップS9で生成された第2クエリグラフパターンの変数ノード「?keyword」にユーザにより入力されたキーワード「セマンティックWeb」を代入し、この第2クエリグラフパターンを基に、共通クエリグラフパターン生成部12からグラフ構造データベース30にアクセスされる(S13)。

50

【 0 0 5 1 】

そして、グラフ構造データベース 30 においてこのキーワード「セマンティックWeb」が代入された第 2 クエリグラフパターンにマッチする第 2 共通サブグラフが抽出される (S 1 4)。

【 0 0 5 2 】

本実施形態においては抽出された第 2 共通サブグラフは 2 件であり、これはクラス「人」とクラス「技術」との間に共通して存在する強い関連性のある情報が、キーワード「セマンティックWeb」に関しては 2 件存在するとともに、この 2 件の中のノード「Person:」に示される「山田太郎」と「田中一郎」とは、キーワード「セマンティックWeb」に関して類似した情報を持つことを示している。

10

【 0 0 5 3 】

抽出された 2 件の第 2 共通サブグラフは、共通クエリグラフパターン生成部 1 2 において、同一のノードを 1 つのノードとして表すように 2 つのグラフがマージされる (S 1 5)。

【 0 0 5 4 】

図 1 8 は、2 件の第 2 共通サブグラフがマージされた状態を示す説明図である。

【 0 0 5 5 】

さらに、このマージされた第 2 共通サブグラフは、出力用 G U I 生成部 1 4 において図 1 9 に示すような表形式の出力用 G U I が生成され (S 1 6)、ユーザ端末 2 0 において表示される (S 1 7)。この際、グラフ構造データベースのデータよりドキュメントの題名がわかっているものについては併せて表示される。図 1 9 においては、ドキュメント「rm:T-041」とともにこのタイトル「セマンティックWeb入門」が表示されている。

20

【 0 0 5 6 】

以上の実施形態によれば、グラフ構造を持つ大量のデータの中からクエリグラフパターンが一致する情報を検索する際に、ユーザにより入力されたキーワード「セマンティックWeb」および概念「人」に関連し意味的に共通の構造を持つ情報を自動的に取得し、ユーザに見やすい形式で提供することができる。

【 0 0 5 7 】

また、ステップ 1 1 において共通クエリグラフパターンが蓄積されることにより、グラフ構造データベース 30 にデータが追加された際にも、蓄積されている共通クエリグラフパターンを基にグラフ構造データベース 30 を再度検索すれば追加されたデータも含まれた情報を取得することができる。

30

【 0 0 5 8 】

本実施形態では、ステップ S 5 において検索クエリグラフパターンを生成する際、2 点のノード間のノードが 1 個の検索サブクエリを検索するための検索クエリグラフパターンを生成しているが、これには限定されず、2 個以上のノードが存在する検索クエリグラフパターンを生成することも可能である。また、キーワードにつながるプロパティによって制約されるノードのみ、もしくはクラスによって制約されるノードのみで 2 点を構成して検索クエリグラフパターンを生成することも可能である。さらには、プロパティの矢印の向きに応じて複数の検索クエリグラフパターンを生成することも可能である。

40

【 0 0 5 9 】

また、本実施形態では、ステップ S 8 において部分第 1 検索サブグラフを抽出する際、2 点のノード間が 2 個以上の複数のパスで構成されていることを条件としているが、これには限定されず、このパスの個数については可変とすることができる。その際、検索結果の中にどれだけの個数の同じ構造があれば抽出するかといった条件についても可変とすることができる。

【 0 0 6 0 】

また、本実施形態では、ステップ S 9 において部分第 1 検索サブグラフに含まれるノードを変数にすることにより共通クエリグラフパターンを生成したが、この方法には限定されず、一部のプロパティ名を変数とした共通クエリグラフパターンを生成することも可能

50

である。

【図面の簡単な説明】

【0061】

【図1】本発明の一実施形態による共通クエリグラフパターン生成装置が用いられている検索システム1の構成を示す全体図である。

【図2】本発明の一実施形態による共通クエリグラフパターン生成装置に接続されているグラフ構造データベース内のデータを示す説明図である。

【図3】本発明の一実施形態による共通クエリグラフパターン生成装置に接続されているグラフ構造データベース内のデータの一部であり、元のリレーショナルデータベースにおけるデータと、このデータをXML形式およびグラフ形式を用いてRDFで表現した状態を示す説明図である。

10

【図4】本発明の一実施形態による共通クエリグラフパターン生成装置に接続されているオントロジデータベース内のデータの一部を示す説明図である。

【図5】本発明の一実施形態による共通クエリグラフパターン生成装置が用いられている検索システムの動作を示すシーケンス図である。

【図6】本発明の一実施形態による共通クエリグラフパターン生成装置に入力される入力GUIを示す説明図である。

【図7】本発明の一実施形態による共通クエリグラフパターン生成装置で取得された、入力キーワードにつながるプロパティを太線で示したグラフ構造データを示す説明図である。

20

【図8】本発明の一実施形態による共通クエリグラフパターン生成装置で生成された検索クエリグラフパターンを示す説明図である。

【図9】本発明の一実施形態による共通クエリグラフパターン生成装置で取得された検索クエリグラフパターンにマッチした検索サブグラフの例を示す説明図である。

【図10】本発明の一実施形態による共通クエリグラフパターン生成装置で取得された検索サブグラフにおいて共通する構造およびプロパティを太線で示した状態を示す説明図である。

【図11】本発明の一実施形態による共通クエリグラフパターン生成装置で生成された共通クエリグラフパターンの一例を示す説明図である。

【図12】本発明の一実施形態による共通クエリグラフパターン生成装置で取得された検索クエリグラフパターンにマッチした検索サブグラフの例を示す説明図である。

30

【図13】本発明の一実施形態による共通クエリグラフパターン生成装置で取得された検索サブグラフにおいて共通する構造およびプロパティを太線で示した状態を示す説明図である。

【図14】本発明の一実施形態による共通クエリグラフパターン生成装置で生成された共通クエリグラフパターンの一例を示す説明図である。

【図15】本発明の一実施形態による共通クエリグラフパターン生成装置で生成された共通クエリグラフパターンにユーザにより入力されたキーワードを代入した状態の一例を示す説明図である。

【図16】本発明の一実施形態による共通クエリグラフパターン生成装置で取得された共通クエリグラフパターンにマッチした共通サブグラフの一例を示す説明図である。

40

【図17】本発明の一実施形態による共通クエリグラフパターン生成装置で生成された共通クエリグラフパターンにユーザにより入力されたキーワードを代入した状態の一例を示す説明図である。

【図18】本発明の一実施形態による共通クエリグラフパターン生成装置で取得された共通クエリグラフパターンにマッチした共通サブグラフの一例を示す説明図である。

【図19】本発明の一実施形態による共通クエリグラフパターン生成装置に接続されているユーザ端末で表示された表形式のデータを示す説明図である。

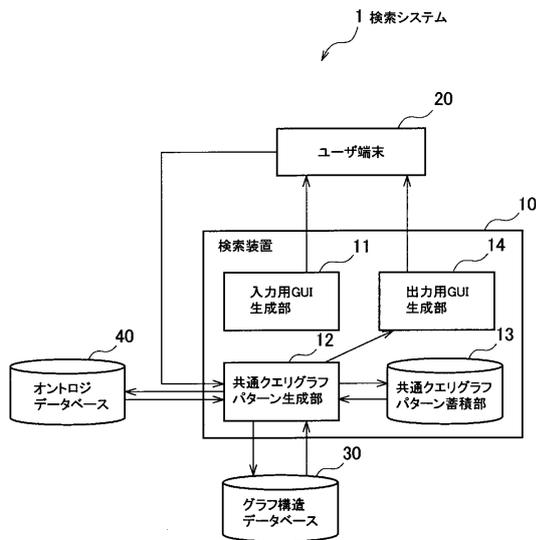
【符号の説明】

【0062】

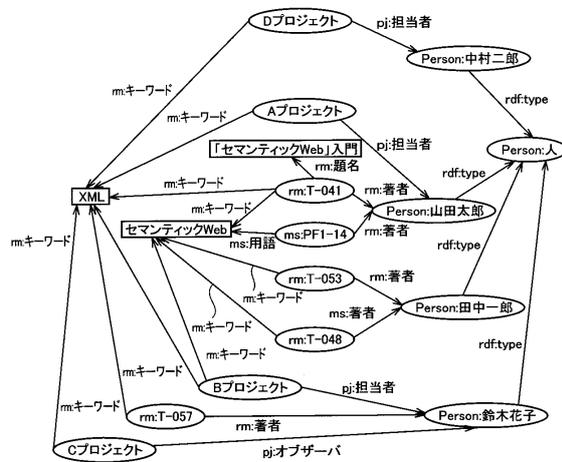
50

- 1 ... 検索システム
- 1 0 ... 検索装置
- 1 1 ... 入力用GUI生成部
- 1 2 ... 共通クエリグラフパターン生成部
- 1 3 ... 共通クエリグラフパターン蓄積部
- 1 4 ... 出力用GUI生成部
- 2 0 ... ユーザ端末
- 3 0 ... グラフ構造データベース
- 4 0 ... オントロジデータベース

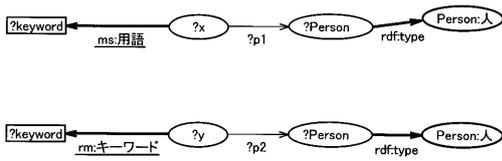
【図1】



【図2】

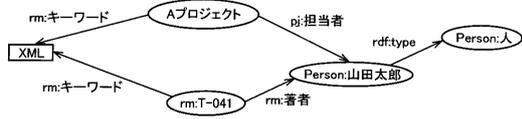


【 図 8 】

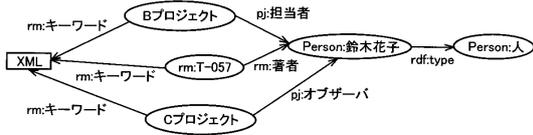


【 図 9 】

(1) 変数keywordの値がXML、変数personの値がPerson:山田太郎となった検索結果



(2) 変数keywordの値がXML、変数personの値がPerson:鈴木花子となった検索結果

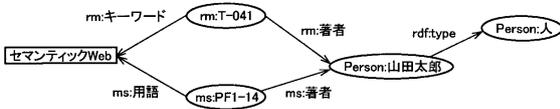


(3) 変数keywordの値がXML、変数personの値がPerson:中村二郎となった検索結果

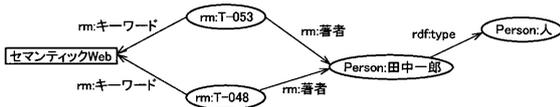


【 図 1 2 】

(1) 変数keywordの値がセマンティックWeb、変数personの値がPerson:山田太郎となった検索結果



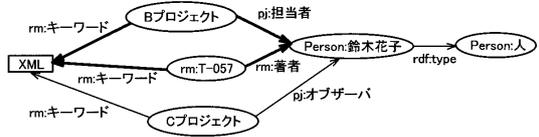
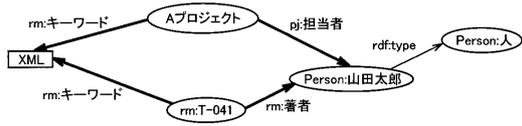
(2) 変数keywordの値がセマンティックWeb、変数personの値がPerson:田中一郎となった検索結果



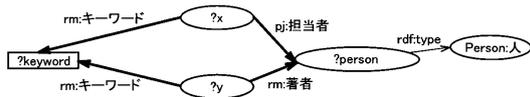
(3) 変数keywordの値がセマンティックWeb、変数personの値がPerson:鈴木花子となった検索結果



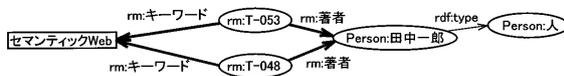
【 図 1 0 】



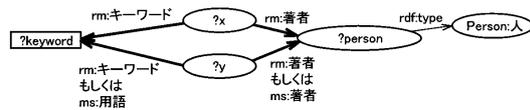
【 図 1 1 】



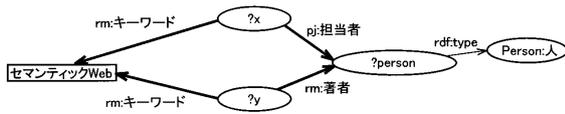
【 図 1 3 】



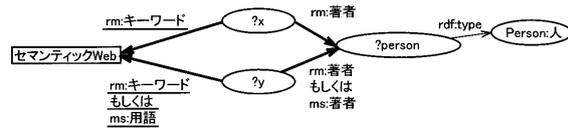
【 図 1 4 】



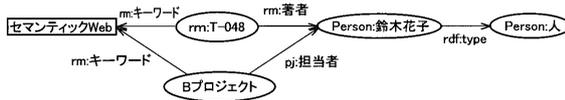
【図15】



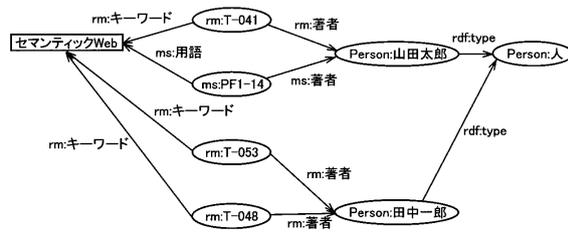
【図17】



【図16】



【図18】



【図19】

セマンティックWeb	←キーワード	rm:T-041 セマンティックWeb入門	著者→	人 山田太郎
	←用語	ms:PF1-14	著者→	
	←キーワード	rm:T-053	著者→	人 田中一郎
	←キーワード	rm:T-048	著者→	

フロントページの続き

(72)発明者 村山 隆彦

東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内

審査官 鈴木 和樹

(56)参考文献 特開2003-099469(JP,A)

特開2003-228580(JP,A)

特開平07-225771(JP,A)

上田俊夫、外3名、研究のツールボックス【1】Jena - セマンティックWebアプリケーション開発のためのJavaフレームワーク, 人工知能学会誌, 日本, (社)人工知能学会, 2004年 5月 1日, 第19巻, 第3号, p. 325 - 333

飯塚京士、外3名、RDFデータを対象としたグラフ検索におけるクエリ生成方式の検討, 第11回セマンティックウェブとオントロジー研究会資料(SIG-SWO-A502-08)[online], 2005年11月16日, p. 1 - 8, URL, <http://www.jaist.ac.jp/ks/labs/kbs-lab/sig-swo/papers/SIG-SWO-A502/SIG-SWO-A502-08.pdf>

(58)調査した分野(Int.Cl., DB名)

G06F 17/30

G06F 13/00

G06Q 10/00 - 50/00